Characterization of TSG101 Induced Allostery within Glucocorticoid Receptor and Computational Drug-Lead Targeting of the Glucocorticoid Receptor DNA-Binding Domain

by

Jordan Thomas White

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

November, 2017

© Jordan Thomas White 2017

All rights reserved

Abstract

Glucocorticoid Receptor (GR) is a typical steroid hormone receptor. GR has a disordered N-terminal domain that binds transcriptional cofactors, a DNA-binding domain, and a steroid binding domain. Most work for the past forty years has been focused on the latter two domains that are structured and amenable to characterization. Yet, the disordered N-terminus makes up more than half of the protein and is absolutely required for transcriptional activation (repression can occur regardless). There are also translational isoforms with truncated N-termini that alter the resulting activity of GR. How the disordered isoforms of the N-terminus couple to cofactors and transcriptional activity is currently unclear. In this thesis, I sought organizing principles to explain the varying activities of GR translational isoforms. I found that a transcriptional cofactor, tumor susceptibility gene 101 (TSG101), can bind and fold the disordered N-terminus of GR. TSG101's binding is allosterically coupled to binding of DNA, but the exact mechanism differs between different isoforms of GR. In this thesis I

also characterized the thermodynamic stability of the TSG101 coiled-coil, which binds GR, and I used computational docking to develop novel drug-leads targeted against the GR DNA-binding domain.

Dedication

$$\mathcal{P}_i = \frac{w_i}{\sum_{j=0}^{n} w_j}$$

where w is a statistical weight and $\mathcal{P}$ is probability.

Acknowledgments

Certainly, I would not be here without the unending support of my family and friends. I am privileged in this regard and grateful. There are a great many people whom I could acknowledge and I admit this is an incomplete filing. There are many people in the Hopkins community that have helped develop me and the thesis contained herein. Brad Thompson gave Prof Hilser and me medical context and sage advice. Jing Li helped me get started in the lab. Dima Toptygin and Arne Shön are perpetual fonts of knowledge. Professors Brand and Moudrianakis have taught me with their wisdom of experience and both have helped me avoid potential experimental pitfalls. My many labmates for thoughtful suggestions. James Rives and Kyle Hoban for pre-reading my thesis. And of course, Prof Hilser and my thesis committee (Dean Wendland & Prof Freire) for years of advice and encouragement. Special thanks to Prof Gray who graciously agreed to join the committee as the secondary reader.

Contents

Abstract **ii**

Acknowledgments **v**

List of Tables **xii**

List of Figures **xiv**

List of Symbols **xvi**

1 Introduction to Nuclear Hormone Receptor Biology **1**

1.1 Abstract 2

1.2 Dynamics and Intrinsic Disorder in NHR's 3

1.3 Inter- and Intra-domain Coupling in NHR's 7

1.3.1 DNA response elements as allosteric effectors 7

1.3.2 LBD ligands as allosteric effectors 8

1.3.3 Splicing and translational isoforms modulate allosteric communication 10
1.4 The Ensemble Allosteric Model: Application in the Case of NHR's . 12
1.5 The Ensemble Allosteric Model may Reconcile Puzzling Observations in NHR's 19
1.6 Conclusions 24

2 Review of the Endosomal Sorting Complex Required for Transport (ESCRT) 26

2.1 Abstract 27
2.2 Endosomal Membranes 28
2.3 Endosomal Maturation 31
2.3.1 ESCRT-0: Signal Recognition 34
2.3.2 ESCRT-I: Signal Transduction 36
2.3.3 ESCRT-II: Signal Amplification 37
2.3.4 ESCRT-III: Wrapping Things Up 39
2.3.5 VPS4 and VTA1: Resolution of Exvaginating Vesicles 41
2.3.6 ALIX 42
2.4 Non-Endosomal Functions of ESCRTs 43
2.5 Disease Implications 46
2.6 Conclusion 47

3 Structural Stability of the coiled-coil domain of Tumor Susceptibility Gene (TSG)-101 48

3.1 Abstract 49

3.2 Introduction 50

3.3 Methods 53

3.3.1 Protein Expression and Purification 53

3.3.2 Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI) 56

3.3.3 CD 56

3.3.4 DSC 57

3.3.5 Fluorescence 57

3.3.6 AUC 59

3.3.7 COREX/BEST Calculations 60

3.3.8 DSC and CD Fitting Equations 60

3.3.9 Fluorescence Methods 65

3.4 Results 67

3.4.1 MALDI mass spectrometry 67

3.4.2 Circular Dichroism (CD) Reveals α-Helical Structure 70

3.4.3 Differential Scanning Calorimetry (DSC) Comports with CD Data 74

3.4.4 Fluorescence Indicates that Pyrene Labeling of TSG101cc Does not Disturb its Structure Significantly . . . 75
3.4.5 Analytical Ultracentrifugation (AUC) Confirms the Two TSG101cc States . . . 83
3.4.6 Structural Thermodynamic Analysis Using COREX/BEST . . 85
3.5 Discussion and Conclusions . . . 96
3.5.1 Discussion . . . 96
3.5.2 Conclusions . . . 99
3.5.3 Acknowledgments: . . . 102
3.6 Appendix . . . 102
3.6.1 Sequences of TSG101cc constructs used . . . 102
3.6.2 PMIA labeling of cysteine . . . 103

4 The Tumor Susceptibility Gene-101 Coiled-Coil Binds and Allosterically Regulates the Glucocorticoid Receptor 104

4.1 Abstract . . . 105
4.2 Introduction . . . 106
4.3 Methods . . . 112
4.3.1 Cloning of Plasmids Expressing Human GR or TSG101 . . . 112
4.3.2 Protein Expression and Purification . . . 114
4.3.3 Protease Protection, Cross-Linking, and Mass Spectrometry 117
4.3.4 Fluorescence Anisotropy . . . 121

4.3.5 Circular Dichroism (CD) 126
4.3.6 Single Molecule Force Spectroscopy 126
4.3.7 Cell Culture 127
4.3.7.1 Luciferase Assays 128
4.3.7.2 Confocal Microscopy 128
4.3.7.3 W-blots 130
4.4 Results 132
4.4.1 TSG101cc Binds the NTD of GR 132
4.4.1.1 Protease Protection 132
4.4.1.2 Cross-linking 137
4.4.2 Binding of the GR NTD and TSG101cc is Coupled to Conformational Change 138
4.4.3 TSG101cc Allosterically Promotes DNA-Binding of GR . . . 147
4.4.4 TSG101 Localizes to the Nuclei of U2OS Cells 150
4.4.5 TSG101 Can Increase the Transcriptional Activity of GR . . 156
4.5 Discussion 161
4.6 Appendix 168
4.6.1 Tables of MS Ions Observed 168
4.6.2 TSG101 Evolutionary Alignment 173
4.6.3 Contour Length Analysis for C3-NTD+TSG101cc 177
4.6.4 Raw Data of 6FAM-DNA Fluorescence Anisotropy 179

4.6.5 Nucleotide and Amino Acid Sequences of Plasmids Cloned in This Work . . . 180
4.6.5.1 Bacterial Expression Construct (pJ411 plasmid) . . 180
4.6.5.2 Mammalian Expression Constructs (pJ603 plasmid) 182
4.6.6 Python Code . . . 187

5 Identification of potential GR drug-leads 197

5.1 Introduction . . . 197
5.2 Methods . . . 200
5.2.1 Luciferase assays . . . 200
5.2.2 Cloning of constructs . . . 202
5.2.3 Immunofluorescence . . . 202
5.2.4 Docking . . . 203
5.2.5 Alignment of the steroid hormone receptors (SHR's) . . . 204
5.3 Results . . . 205
5.3.1 Verification of the allosteric site . . . 205
5.3.2 In silico drug-lead screening . . . 209
5.4 Conclusions . . . 215
5.5 Appendix . . . 217

Bibliography 232

Vita 309

List of Tables

1.1 Breakdown of Allosteric States and Energies 15

2.1 List of ESCRT proteins 33

3.1 Fitted Thermodynamic Values for CD 72
3.2 Fitted Thermodynamic Values for DSC of TSG101cc 75
3.3 Fitted Correlation Time Values 80
3.4 Fitted Intensity Decay Values 80
3.5 Structural Thermodynamic Analysis of the TSG101cc 90
3.6 Structural Thermodynamic Analysis of Yeast ESCRT-I 91
3.7 Review of Thermodynamic Data for Oligomeric Proteins in the Literature 93
3.8 Apolar Surface Area Comparison of Oligomeric Proteins in the Literature 94
3.9 Continuation of Prior Two Tables 95

4.1 Compilation of Literature Data on TSG101's Transcriptional Effects and Interactions 111
4.2 Compilation of SMFS Results 145
4.3 Table of Fitted Binding Constants 149
4.4 Ions Detected in MS of Protease Protection 168
4.5 Ions Detected in MS of Unreacted C3NTD and TSG101cc 169
4.6 Ions Consistent with GR or TSG101 in MS of Reacted C3NTD and TSG101cc #1 170
4.7 Ions Consistent with GR or TSG101 in MS of Reacted C3NTD and TSG101cc #2 171
4.8 TSG101cc X-links 171
4.9 Ions of C3NTD and TSG101cc Observed After X-linking 172

5.1 Chemical Make Up of the Drug-Lead Clusters 212

5.2 List of Clean Drug-Leads 218
5.3 List of Fragment Drug-Leads 223

List of Figures

1.1 The Domain Organization of NHR's 4
1.2 Cartoon of an Allosterically Coupled Protein with Three Subunits . 13
1.3 Examples of CR for Single Ligand Binding 17
1.4 Positive-Negative Response Switching 18

2.1 An Overview of Endosomal Maturation 28
2.2 Activation of EGFR 29
2.3 Cartoon Schematic of ESCRT-0—II 32
2.4 Speculative Cartoon of ESCRT-III Vesicle Formation 44

3.1 Schematic of TSG101 Domain Organization 52
3.2 SDS-PAGE gel of SEC purified TSG101cc 54
3.3 MALDI Data Verifying Identity of TSG101cc 69
3.4 Circular Dichroism Wavelength Scan of a Temperature Melt of TSG101cc 70
3.5 Singular Value Decomposition of Circular Dichroism Melt Data as a Function of Wavelenth 71
3.6 Circular Dichroism Thermal Melt of TSG101cc 73
3.7 Differential Scanning Calorimetry of TSG101cc 76
3.8 Steady State Fluorescence of TSG101cc-pyrene 78
3.9 Polarized TCSPC Data for TSG101cc-pyrene 79
3.10 Decay-Associated Amplitude Spectra of TSG101cc-pyrene 81
3.11 AUC of the TSG101cc 84
3.12 Structural Thermodynamic Analysis of TSG101cc 86
3.13 The Energy of ΔS_{conf} 88
3.14 Concluding Graphic 101

4.1 The Domains and Functions of GR and TSG101 110
4.2 GR and TSG101 Protein Purity 118
4.3 TSG101cc Protects the N-terminus of GR from Protease Digestion 134

4.4 Control Trypsin Digests . . . 135
4.5 Mass Spectrometry of Protease Protection Assays . . . 136
4.6 Cross-Linking of C3-NTD and TSG101cc . . . 139
4.7 Cross-Linking Results and Speculative Model . . . 140
4.8 CD Reveals a Structural Change in Binding of TSG101cc and the C3-NTD . . . 141
4.9 CD Reveals a Structural Change in Binding of TSG101cc and the C3-NTD . . . 143
4.10 TSG101cc can Fold the N-terminus of GR . . . 144
4.11 Transitions Seen in SMFS Depend Upon the TSG101cc Binding Site of GR . . . 146
4.12 Quantitative Binding of TSG101cc to GR . . . 148
4.13 TSG101 Localizes to U2OS Nuclei . . . 152
4.14 TSG101-tdTom Localizes to U2OS Nuclei . . . 153
4.15 Overexpression of TSG101 for 16 Hours . . . 154
4.16 Overexpression of TSG101 Eventually Causes Endosomal Enlargement . . . 155
4.17 Localization of GR and TSG101 . . . 156
4.18 TSG101 Activates GR . . . 158
4.19 TSG101's Coiled-Coil is Necessary for Activation . . . 159
4.20 TSG101 Stabilizes a Transcriptionally Active Form of GR . . . 160
4.21 TSG101 May Directly Activate GR . . . 167
4.22 Amino Acid Alignment of Human TSG101 and Ten Homologues . . . 173
4.23 Compilation of C3-NTD Force Ramp Data . . . 177
4.24 Compilation of $AF1^{107\text{-}237}$ Force Ramp Data . . . 178
4.25 TSG101cc Does Not Directly Bind DNA . . . 179

5.1 The Domain Organization of GR . . . 198
5.2 Verification of the DBD Allosteric Site . . . 206
5.3 Confocal Immunofluorescence of the R-NLS-FLAG Construct . . . 208
5.4 Luciferase Assay of C3-DBD with the R-NLS-FLAG Construct . . . 209
5.5 The Allosteric Surface of the GR DBD . . . 210
5.6 Structures of Top Hits . . . 211
5.7 Clustering of the Top 105 Docked Molecules . . . 213
5.8 Alignment of other Steroid Hormone Receptors to GR . . . 214
5.9 Comparison of GR Docking Scores Versus Other SHR's . . . 216
5.10 Comparison of GR and PR Docking . . . 217
5.11 Top Two Fragment Drug-Leads . . . 223

List of Symbols

- SHR = Steroid Hormone Receptor $\subset$ NHR = Nuclear Hormone Receptor (SHR's are a subset of NHR's)
- GR = Glucocorticoid Receptor; GRE = GR Response Element (the DNA GR binds as a dimer)
- NTD = N-Terminal disordered Domain of a SHR (GR in this thesis)
- DBD = DNA-Binding Domain of GR
- LBD = Ligand-Binding Domain of GR
- NTD-DBD = a construct using just the first two domains of GR
- TSG101 = Tumor Susceptibility Gene-101
- TSG101cc = TSG101's coiled-coil domain

Chapter 1

Introduction to Nuclear Hormone Receptor Biology

Publication note: This chapter is largely a republication of a book chapter that the author wrote with several others ([437]). My co-authors are: Hesam Motlagh, Jing Li, E. Brad Thompson, and Vincent J. Hilser.

1.1 Abstract

Nuclear Hormone Receptors (NHRs) must respond to a variety of chemical cues to selectively induce or repress pertinent genes. Each domain of these multidomain proteins is capable of binding to various large or small molecules, and such binding may alter the activity of both the local and remote domains. These allosteric effects are essential for proper NHR function; yet until recently very little was known about their mechanism. Recent well-documented findings show that protein dynamics and intrinsic disorder mediate allosteric signals. I discuss the recently proposed Ensemble Allosteric Model, which can be used to mechanistically dissect perplexing and sometimes paradoxical allosteric phenomena. Finally, I close with how this model can be specifically applied to NHRs and how its use can enhance future studies.

Physiological and molecular biological studies have dominated nuclear hormone receptor (NHR) research for the past half century. Many experiments have demonstrated that NHRs function as allosteric proteins. Recent thermodynamic studies have begun to show how NHRs modulate their biological response by using allostery, wherein an energetic perturbation at one site mediates a functional response at a physically distinct site [201, 219, 242]. These energetic perturbations can be affected by many physical phenomena, including but not limited to: ligand binding, post-translational modifications, and different NHR

protein isoforms [292, 345, 435]. It has become clear that transcription factors, generally [106, 251], and NHRs specifically [201, 219, 242], utilize intrinsic disorder to modify their allosteric responses. This chapter will first describe dynamics and intrinsic disorder in NHRs, and then show this intrinsic disorder is poised to maximize allosteric coupling in nuclear hormone receptors.

1.2 Dynamics and Intrinsic Disorder in NHR's

One of the most rapidly expanding fields in protein biophysics is the study of intrinsic disorder in proteins [410]. In the past decade, it was realized that these intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are in perhaps two-thirds of eukaryotic transcription factors [251]. As a major class of transcription factors, NHRs are no exception and all NHRs show some disorder propensity [214]. Most NHRs are multidomain proteins consisting of ligand-binding (LBD), DNA-binding (DBD) and sometimes, N-terminal (NTD) domains Figure 1.1.

The LBD contains several dynamic regions that are predominantly near the ligand-binding pocket. Some of the earliest research on the dynamic nature of the LBD is from the O'Malley lab in the 1990's. Working with progesterone and estrogen receptors, they found that the LBD undergoes a conformational change upon binding hormone [9, 9, 40]. Subsequent X-ray crystallography and

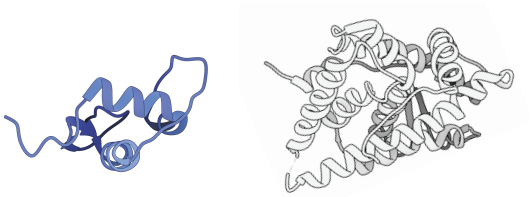


Figure 1.1 – The Domain Organization of NHR's
The diagram reads N-terminus on the left, C-terminus on the right. The N-terminal domain (NTD) is intrinsically disorder and often binds cofactor proteins. The DNA-binding domain (DBD) binds response elements. The ligand-binding domain (LBD) binds small molecules, such as steroids, and some cofactor proteins. The NTD length is highly variable and its sequence is poorly conserved. The DBD and LBD together are on the order of 300 amino acids total and are well conserved. Example structures are shown above the box diagram (GR DBD and LBD are pdb:1glu and 1p93, respectively).

fluorescence anisotropy experiments demonstrated that when the LBD binds ligand, helix-12 undergoes a conformational change and becomes less dynamic [53, 190, 341, 423]. This conformational change varies depending on the ligand bound; hence different ligands can promote specific conformations that bind corepressors or coactivators [59, 298]. Researchers have also shown that cofactor fragments that bind the LBD can stabilize the LBD interaction with certain ligands [119, 190], demonstrating an allosteric communication between cofactor and ligand binding. The cofactor thus acts as an allosteric effector/ligand.

The DBD also has regions that are dynamic. Estrogen and glucocorticoid receptor DBDs are partially disordered in solution [43, 363], and become more ordered when bound to DNA [263, 362]. Spolar and Record first described this transition on thermodynamic grounds [379], and their analysis hinted at the

possibility of different DNA sequences having different effects on the DBD. Indeed, the sequence of DNA does modulate this conformational change [30, 257, 278]. Specifically, the DNA sequence can change both the binding affinity to the DBD and the propensity for dimerization of the DBD [126, 163, 444]. It has been speculated that response elements cause allosteric effects beyond the DBD—possibly changing the conformation of the NTD [238, 383], a point that will be revisited in the next section.

The steroid hormone receptor (SHR) subclass of NHRs contains large NTDs of varying size and sequence, which are perhaps the best example of dynamics and disorder in the NHR family of proteins [214]. Each type of SHR contains a large amount of ID in the N-terminus, as determined by a myriad of biophysical methods [28, 29, 79, 108, 218, 230, 304]. It is not yet clear why SHR N-termini are disordered, but folding of these domains appears to be key to understanding their function. Many disordered proteins undergo coupled folding and binding to perform their biological role [417], and in 1999, the Thompson lab presented the first evidence that the NTD of an SHR could fold into a tertiary structure, either because of the DBD binding a response element or because of high concentrations of an osmolyte [38, 215]. When so folded, the NTD showed enhanced binding of known partner proteins [216]; thus, the folded state of an SHR NTD is likely the biologically functional state. Since then, many groups have used various osmolytes, such as tri-methyl amine N-oxide (TMAO; see [48]) to fold the NTDs

of NHRs and have demonstrated that they fold cooperatively, a hallmark of a naturally evolved, folded state [433].

Glucocorticoid, androgen, and mineralocorticoid receptors' NTDs all fold cooperatively in the presence of molar concentrations of TMAO [38, 108, 242, 338]. This is evidence of a major folded conformation or an ensemble of folded conformations. As mentioned above, the folded state is the presumed biologically active state of the NTD because of studies showing that the NTD folds upon binding transcriptional cofactors. The NTDs of SHRs bind numerous cofactor proteins [255], and studies on estrogen, glucocorticoid, and androgen receptors have shown that the NTDs become more ordered upon binding cofactor proteins [75, 201, 219, 338, 430]. This suggests a coupled folding and binding mechanism regulates SHR activity.

In sum, all NHR domains—even the relatively structured LBDs and DBDs—are dynamic and exist as conformational ensembles of states. Upon binding a ligand, domains are stabilized in one or another globular state of a more limited ensemble. These ligand-specific effects have functional consequences on the co-regulators bound, and on transcriptional function. By varying the transcriptional function of a receptor, ligands can alter cellular and clinical outcomes. "Ligand" thus refers to all binding partners—protein, DNA, steroidal or other small molecules—as they interact with their respective domains.

1.3 Inter- and Intra-domain Coupling in NHR's

Binding of a ligand in one NHR domain not only affects the conformational ensemble of that domain, but also the conformational ensembles of the other domains. This allosteric coupling between domains is crucial for the function of NHR's.

1.3.1 DNA response elements as allosteric effectors

Early glucocorticoid research raised a fundamental question [175]—how does a NHR activate a multitude of genes to different degrees? One possibility is that each response element could modify the activity of a bound NHR. This idea was supported by research on glucocorticoid, estrogen, and thyroid hormone receptors [9, 126, 238, 269, 353, 383]. In particular, work from Yamamoto and colleagues demonstrated that glucocorticoid receptor activity was dependent on the response element sequence [353] and also speculated that a conformational change was occurring upon DNA binding.

Ikeda and colleagues provided the first evidence of different response elements inducing different conformational changes in NHRs [167]. They found that transcriptionally active response elements induced a change in thyroid hormone/retinoid X receptor dimers, such that the complex was resistant to protease digestion, relative to protein dimers on inactive response elements [167]. Later

studies found similar effects with estrogen and androgen response elements and their cognate receptors [122, 257, 443, 444].

The limited proteolysis experiments described above demonstrate DNA response element dependent protection patterns. However, which portions of the molecule were undergoing conformational changes was still an open question. In 1999, Kumar and colleagues published the first manuscript demonstrating that when an NHR, glucocorticoid receptor, binds DNA its NTD undergoes a conformational change [215]. Kumar used a combination of circular dichroism and tryptophan fluorescence to demonstrate that folding was occurring in the NTD. Subsequent studies have revealed similar phenomena in progesterone receptor [28, 29], estrogen receptor alpha [131], and androgen receptor [58]. The inferred significance of these observations is that binding of DNA coupled to folding of the NTD would recruit transcriptional cofactors, as proposed by Thompson and coworkers [75, 403].

1.3.2 LBD ligands as allosteric effectors

Different ligands binding to the LBD can elicit specific transcriptional responses. Early results on this matter are conflicting. Several studies suggested that binding of hormone increased the binding affinity for DNA [26, 39, 450] while others did not [206, 436]. Some of these ambiguities are likely due to different DNA response elements used by different labs. As shown by Meyer and coworkers,

different response elements can modulate the effect of a given hormone. They demonstrated that RU486, a known antagonist, of the progesterone receptor A isoform could activate the B isoform. Furthermore, this activation only occurred on one of two promoter sequences that they tested [281]. This is an example of selective response modification by an NHR, in which a ligand modifies receptor activity in a manner dependent on the DNA bound [121]. By this and probably other mechanisms, selective sets of genes are activated or repressed by specific steroids acting through their cognate receptors. Such ligands activate distinct, but usually overlapping sets of genes.

Selective response modifiers exert their effects through a number of mechanisms. Besides affecting DNA binding, ligands also change subsequent events. Different ligands alter recruitment of NTD binding partners to estrogen and glucocorticoid receptors [118, 345, 366]. Furthermore, ligands can also alter dimerization of NHRs, as seen in the formation of PPARα-RXRα heterodimers on DNA [109]. Regardless of the variety of mechanisms, what is clear from these examples is that ligand binding at the LBD causes allosteric effects on the DBD and NTD.

1.3.3 Splicing and translational isoforms modulate allosteric communication

Expression of tissue-specific protein isoforms has the potential to dictate what makes a liver a liver or a heart a heart, and recent work has shown that tissue-specific coding exons are enriched in intrinsic disorder and protein binding motifs [61]. Interestingly, NTDs of NHRs are often intrinsically disordered, full of co-regulator binding sites, and can exist in multiple isoforms [69, 214, 260, 359, 399, 411]. Several studies have also demonstrated that NHR isoforms differ in their transcriptional activities [260, 281, 292]. However, it is rarely made clear how isoforms have different transcriptional activities or specificities. One possibility is that each isoform contains a unique system of allosteric coupling and disorder. If so, this raises the possibility that tissues tune the activity of a NHR by expressing varying amounts of its NTD isoforms.

Three steroid receptors are known to possess alternative NTD isoforms (progesterone, glucocorticoid, and ecdysone receptors; see: [196, 260, 399]). Of the three ecdysone receptor isoforms, two have very similar NTD lengths and yet vastly different effects on transcriptional activity [88, 292]. Similar effects have been reported for progesterone and glucocorticoid receptors [260, 281], and in the case of glucocorticoid receptor this change in NTD length has an effect on the folding free energy of this disordered domain [242]. Interestingly, the

isoform with the lowest folding energy is transcriptionally the most active. This is most readily explained by coupled folding of NTD with binding of coregulators because a more stable isoform will bind to coregulators more strongly.

LBD isoforms also occur. For example, the glucocorticoid receptor has three exon splice-variant isoforms that alter the LBD length [350]. Of these, only GRα is able to bind hormone. The other two (β and P) have truncated LBDs and markedly different effects on transcriptional regulation [86, 351]. Mechanistically, the shorter LBDs ablate the ability of hormone to induce a change in the receptor. However, it is unclear how these changes affect LBD allosteric coupling to the other receptor domains.

In sum, NHRs use inter- and intra-domain coupling to govern their responses to a variety of allosteric effector ligands and thus determine the specificity of their function. Binding of its ligand at any given domain (NTD, DBD, LBD) can affect the stability and binding affinity of their ligands at other domains. These allosteric changes have been understood from a phenomenological view for many years. Only recently has a model been developed that explains these phenomena in testable, quantitative terms. Below I will discuss this general model of allostery (the Ensemble Allosteric Model) that can be used to glean mechanistic insight into NHR allostery.

1.4 The Ensemble Allosteric Model: Application in the Case of NHR's

How allostery works is a century-old question [2, 36] that was initially addressed with hemoglobin. Monod, Wyman, and Changeux proposed arguably the most influential allosteric model using macroscopic, thermodynamic concepts to take into account a conformational change within the subunits of hemoglobin [286]. Other influential models have been proposed that explain most of the available data on hemoglobin [144, 211]. However, all these models have limitations as they are phenomenological and do not address "how" allostery is mediated between distal sites [289]. Crystallographic or other structural data has been used to elucidate bond paths linking the binding site of the allosteric effector with a distant responsive site [254, 313–315, 397]. Even though it has been suggested that these allosteric pathways may be dominant [80], they do not explain the following: all of the available data on hemoglobin [375], proteins in which no bond path can be found between the allosteric effector binding site and the distant response region, or the perplexing observation of protein dynamic- and disorder- mediated allostery [106, 316, 337].

The Hilser lab recently proposed a model that alleviates some of these issues by articulating allostery in terms of the intrinsic energetics of a protein – the

Ensemble Allosteric Model (EAM) [150, 289]. Because it applies to all proteins, the EAM can be used to describe allostery in NHRs. In particular, it can explain perplexing phenomena, such as the ability of a single ligand to be an agonist or an antagonist to NHR function in different cells [13, 442].

The EAM is grounded in two well-established observations from protein allostery: The ability of allosteric proteins to exist in multiple conformations in solution, and their segregation of binding sites into different domains. Such domains can communicate with one another, the essence of allostery. From these simple and well-established facts, it is possible to develop the model and ask quantitative questions about allostery.

Consider an allosteric protein consisting of three interacting domains (Figure 1.2). The simplest conformational ensemble of each domain is a two state equilibrium between at least one **high affinity (H)** and **low affinity (L)** state. The L state of a domain can be either intrinsically disordered or an ensemble of somewhat ordered conformers that bind its ligand with lower affinity than does the H conformation. The H and L states for each domain

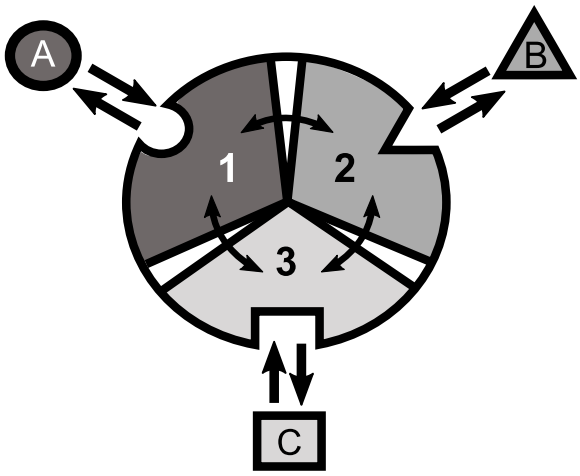


Figure 1.2 – Cartoon of an Allosterically Coupled Protein with Three Subunits
The three subunits (1, 2, 3) bind to one ligand each (A, B, C) and are capable of allosteric communication. In the context of NHR's, the three subunits are the NTD, DBD, and LBD binding to coregulator, DNA, and ligand, respectively. Modified from [289] with permission.

have a free energy difference (i.e. ΔG_i), which determines how often the molecule is in the H or L conformation in the absence of influences from other domains. Because the domains communicate to one another, there must be an interaction energy between them. When one domain goes to its L conformation it either stabilizes (i.e. $\Delta g_{1,2} < 0$) or destabilizes (i.e. $\Delta g_{1,2} > 0$) the H state of the other domain(s) to which it is coupled (Figure 1.2 and Table 1.1).

From this simple articulation, enumeration of all combinations of H and L domain states in addition to their relative free energies is straightforward. Table 1.1 lists every possible combination of domains 1, 2, and 3 being in either the H or L conformation. The free energy of each state is simply the sum of the conformational energy and the interaction energy. Taking the statistical weight of each state from Table 1.1 and summing it, yields the partition function ($\mathcal{Q}$) which gives access to every quantity of interest, including experimental observables:

$$\mathcal{Q} = \sum_{j}^{i} S_j = S_{\mathcal{HHH}} + S_{\mathcal{HHL}} + S_{\mathcal{HLH}} + S_{\mathcal{LHH}} + S_{\mathcal{HLL}} + S_{\mathcal{LLH}} + S_{\mathcal{LHL}} + S_{\mathcal{LLL}} \tag{1.1}$$

With this, it is possible to reproduce basic allosteric phenomena and ask basic questions. For instance, what happens when ligand A is introduced? The H conformation of each domain will preferentially bind its proper ligand at its introduction into the system [445]; thus, introduction of ligand A will stabilize

State	$\Sigma \Delta G_i$	$\Sigma \Delta g_{i,j}$	S_i	Probability
HHH	0	0	1	S_{HHH} / Q
LHH	ΔG_1	$\Delta g_{12} + \Delta g_{13}$	$K_1\varphi_{12}\varphi_{13}$	S_{LHH} / Q
HLH	ΔG_2	$\Delta g_{12} + \Delta g_{23}$	$K_2\varphi_{12}\varphi_{23}$	S_{HLH} / Q
HHL	ΔG_3	$\Delta g_{13} + \Delta g_{23}$	$K_3\varphi_{13}\varphi_{23}$	S_{HHL} / Q
HLL	$\Delta G_2 + \Delta G_3$	$\Delta g_{12} + \Delta g_{13} + \Delta g_{23}$	$K_2K_3\varphi_{12}\varphi_{13}\varphi_{23}$	S_{HLL} / Q
LHL	$\Delta G_1 + \Delta G_3$	$\Delta g_{12} + \Delta g_{13} + \Delta g_{23}$	$K_1K_3\varphi_{12}\varphi_{13}\varphi_{23}$	S_{LHL} / Q
LLH	$\Delta G_1 + \Delta G_2$	$\Delta g_{12} + \Delta g_{13} + \Delta g_{23}$	$K_1K_2\varphi_{12}\varphi_{13}\varphi_{23}$	S_{LLH} / Q
LLL	$\Delta G_1 + \Delta G_2 + \Delta G_3$	$\Delta g_{12} + \Delta g_{13} + \Delta g_{23}$	$K_1K_2K_3\varphi_{12}\varphi_{13}\varphi_{23}$	S_{LLL} / Q

Table 1.1 – Breakdown of Allosteric States and Energies
This table is modified from [289]. ΔG_i is the free energy difference between the high (H) affinity and the low (L) affinity state of domain i. $\Delta g_{i,j}$ is the energy of interaction between two subunits, i and j. S_i is the statistical weight for a given state, where K_i and φ_i are the statistical weights for individual conformations and interactions, respectively. Q is the sum of all statistical weights: $S_i = e^{\frac{\sum \Delta G_i + \sum g_{i,j}}{RT}}$

each microstate in Table 1.1 that has domain 1 in the H conformation by a free energy of:

$$\Delta g_{Lig\mathcal{A}} = -RT * ln(1 + K_a[\mathcal{A}]) = -RT * ln(Z_{Lig,\mathcal{A}}) \tag{1.2}$$

The term $Z_{Lig,\mathcal{A}} = (1 + K_a * [\mathcal{A}])$, is the effect of A on the statistical weights describing domain 1 in the H conformation. Specifically, the partition function in the presence of ligand A is now:

$$Q_{w/\mathcal{A}} = Z_{Lig,\mathcal{A}}(S_{HHH} + S_{HHL} + S_{HLH} + S_{HLL}) + S_{LHH} + S_{LLH} + S_{LHL} + S_{LLL} \tag{1.3}$$

where $Z_{Lig,\mathcal{A}}$ acts as a weighting term that takes into account the increase in the probability of the H conformation of domain 1 in the presence of A. Note that at

$\mathcal{A} = 0$ this equation reduces to the original partition function.

Ligand binding to domain 1 changes the statistical weight of some states, but in turn this will change the probability of all states. Of particular interest is the change in probability of the H state of an "active site" (Domain 1; e.g. the binding site for a coregulator protein in the NTD). The probability of domain 1 being in the high affinity, active state without A present is simply the statistical weight of states where domain 1 is in the H state divided by the partition function:

$$\mathcal{P}_{1,\mathcal{H}}([\mathcal{A}] = 0) = \frac{S_{\mathcal{H}\mathcal{H}\mathcal{H}} + S_{\mathcal{H}\mathcal{L}\mathcal{H}} + S_{\mathcal{L}\mathcal{H}\mathcal{H}} + S_{\mathcal{L}\mathcal{L}\mathcal{H}}}{\mathcal{Q}} \tag{1.4}$$

With A present, all states with domain 1 in the H state will be redistributed:

$$\mathcal{P}_{1,\mathcal{H}}([\mathcal{A}] > 0) = \frac{\mathcal{Z}_{\mathcal{L}ig,\mathcal{A}}(S_{\mathcal{H}\mathcal{H}\mathcal{H}} + S_{\mathcal{H}\mathcal{L}\mathcal{H}}) + S_{\mathcal{L}\mathcal{H}\mathcal{H}} + S_{\mathcal{L}\mathcal{L}\mathcal{H}}}{\mathcal{Q}_{w/\mathcal{A}}} \tag{1.5}$$

Upon binding domain 1, ligand A will redistribute the ensemble and will either increase or decrease the probability of domain 2 being active. To relate the change in probability to the amount of energetic perturbation, I define a value called the Coupling Response (CR) [150] which is the change in probability of a state normalized to the energy of ligand binding:

$$\mathcal{CR}_{2,\mathcal{A}} = \frac{\Delta\mathcal{P}_{2,\mathcal{H}}}{\mathcal{ln}(\mathcal{Z}_{\mathcal{L}ig,\mathcal{A}})} = \frac{\mathcal{P}_{2,\mathcal{H}}([\mathcal{A}] > 0) - \mathcal{P}_{2,\mathcal{H}}([\mathcal{A}] = 0)}{\mathcal{ln}(\mathcal{Z}_{\mathcal{L}ig,\mathcal{A}})} \tag{1.6}$$

A positive CR indicates that ligand A binding to its own domain increases the probability of domain 3 being in the active or H conformation. The opposite is true for negative CR values, which represent negative effects on the stability of a domain (Figure 1.3).

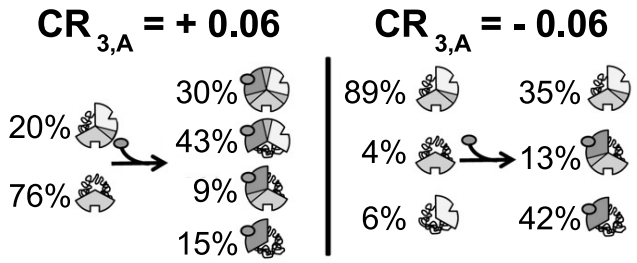


Figure 1.3 – Examples of CR for Single Ligand Binding
A positive $CR_{3,A}$ indicates that binding of ligand A to domain 1 stabilizes states where domain 2 is in its high affinity conformation. Domains 1, 2, and 3 are displayed as the top left, right, and bottom of the circle, respectively. The parameters of the positive $CR_{3,A}$ are: $\Delta G_1 = -1.7$, $\Delta G_2 = 2.0$, $\Delta G_3 = -0.9$, $\Delta g_{12} = -2.3$, $\Delta g_{23} = 0.1$, $\Delta g_{13} = 1.5$, and $\Delta g_{Lig,A} = -5.0$ in kcal/mol. The parameters of the negative $CR_{3,A}$ are: $\Delta G_1 = -2.1$, $\Delta G_2 = 1.0$, $\Delta G_3 = 1.2$, $\Delta g_{12} = -1.7$, $\Delta g_{23} = 0.6$, $\Delta g_{13} = -2.7$, and $\Delta g_{Lig,A} = -5.0$ in kcal/mol. From [289] with permission.

In this three-domain protein example, one can consider how an additional ligand that binds domain 2 would affect the allosteric response of domain 3. When ligand B binds domain 2 it can change the magnitude of the response of domain 3 to ligand A bound in domain 1. The coupling response of other domains to ligand A:Domain 1 binding can change from positive to negative and visa-versa. In a similar manner as before, one can consider the CR of domain 3 when ligand A is added while ligand B is already present:

$$\mathcal{CR}_{3,\mathcal{A}}(\mathcal{B}>0) = \frac{\mathcal{P}_{3,\mathcal{H}}([\mathcal{A}]>0,\mathcal{B}>0) - \mathcal{P}_{3,\mathcal{H}}([\mathcal{A}]=0,\mathcal{B}>0)}{ln(\mathcal{Z}_{\mathcal{L}ig,\mathcal{A}})} \tag{1.7}$$

Note that this equation takes into account the effect of ligand B by itself, thus the CR being described tells us how ligand B changes the allosteric response to ligand A. With two ligands present one can begin to see that some combinations of parameters exist such that ligand A can act as either a positive or a negative regulator of domain 3, a paradoxical observation that has been noted in the NHR field and is discussed in the following section (Figure 1.4).

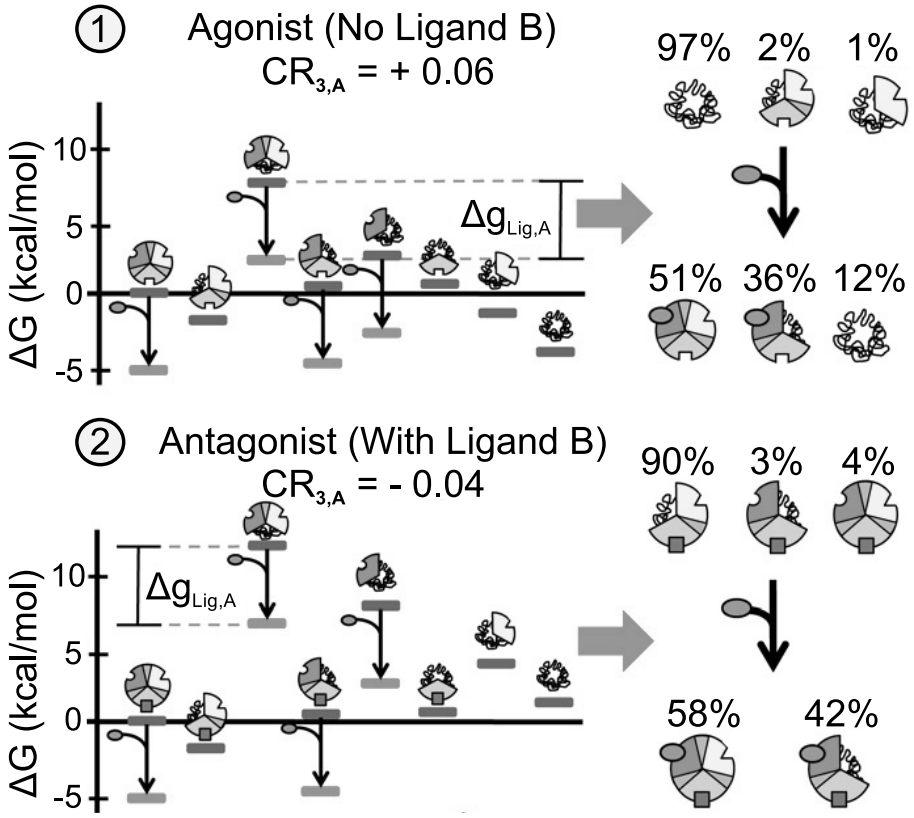


Figure 1.4 – **Positive-Negative Response Switching**
In case 1, ligand A acts as a positive regulator of domain 2. In case 2 the presence of ligand B causes ligand A to act as a negative regulator of domain 2. The parameters used: $\Delta G_1 = -6.75$, $\Delta G_{2,\,B=0}$ (case 1) = − 4.4, $\Delta G_{2,\,B>0}$ (case 2) = 0.6, $\Delta G_3 = -2.7$, $\Delta g_{12} = 6.8$, $\Delta g_{23} = 4.8$, $\Delta g_{13} = -1.9$, and $\Delta g_{Lig,A} = -5.0$ kcal/mol. Domains 1, 2, and 3 are displayed as the top left, right, and bottom of the circle, respectively. From [289] with permission.

1.5 The Ensemble Allosteric Model may Reconcile Puzzling Observations in NHR's

The above model is general and applicable to all allosteric systems, including the three domain SHRs. Each of the three domains of an SHR has binding sites for multiple binding partners, similar to the articulation of the EAM. The C-terminal LBD binds small steroidal and synthetic ligands; it also binds a variety of co-regulatory proteins. The central DBD binds DNA sequences and other proteins; DNA acts as a ligand and can have sequence specific effects on transcription and DBD stability. The NTD also binds co-regulators. Application of the EAM to the SHRs at once makes it apparent that all these binding partners must be considered ligands, as the preceding discussion explains. Furthermore, a simpler version of the EAM obviously applies to the two domain group of NHRs [150].

The model has the potential to explain some perplexing observations in NHR research that have important implications for practical applications and drug development. Tamoxifen is known to inhibit breast cancer and yet promote uterine cancer [442], glucocorticoids exert cell-specific anti-inflammatory effects [13], and a number of other NHR's have effects in a tissue or isoform specific manner [138, 415, 426]. Furthermore, many steroidal and non-steroidal ligands for the LBD act as selective response modifiers (SRMs), meaning that they alter

the transcription of selected, though overlapping, sets of genes, in a cell-specific way.

These phenomena may be understood in terms of thermodynamic ensembles. In different cellular contexts, the equilibria of the ensemble could be tuned by perturbations including but not limited to: (1) type and accessibility of DNA response elements; (2) abundance and type of natural/synthetic ligands to LBD (3) the type and abundance of coregulators to NTD and LBD; (4) different distribution of NHR splicing and translational isoforms; (5) different post-translational modifications to the NHR; (6) effects of intracellular pH, small ion and organic osmolyte levels. All of these will affect the energetic landscape of the ensemble.

As mentioned previously in this chapter, there are known allosteric effects of DNA sequence on the DBD and on NHR function [28, 167, 278, 444]. Binding to different gene elements may bias an NHR to bind certain coregulators. This represents a ligand-based effect on the thermodynamics of another domain. Since it has been shown for glucocorticoid and progesterone receptors that DBD binding of DNA results in acquisition of structure and function in the NTD, there is physical evidence to support this allosteric effect [28, 215, 218].

Recent results suggest that the response element itself is part of the concentration limiting step of transcriptional induction [46], thus gene elements may be a sensitive area of regulation. Gene elements could vary in accessibility because of different DNA methylation patterns in tissues (for review [116]), expression

of different histones [136], or different post-translational modifications of histones [136]. NHR actions do appear to be linked to DNA methylation and to histone modifications [161, 193, 300], and how this affects NHR function is still an active area of research. Put together, there is a large repertoire of perturbations that can modulate the energetics of DNA binding and thus the macroscopic, biological effect.

Nature may also produce different NHR activities through LBD binding of structural variants of hormones [90, 305]. Hormone:LBD binding affects both LBD and NTD stability and their subsequent binding of co-regulators. While the major steroid-producing glands are the source of most circulating steroid hormones, it is now clear that local, tissue-specific steroid synthesis and metabolism can cause the local concentration and type of steroid ligand to vary dramatically. In addition, cell and tissue variations of SHR isoforms and concentrations can vary. The concentration of SHR in a cell can shift the dose-response curve to its cognate ligand by up to an order of magnitude [395, 396]. All of these different hormonal effects would change the probabilities in the conformational ensemble of the LBD. This in turn would affect the distribution of states for other domains of a given NHR.

Binding of different coregulators is a third way to vary NHR activity. There are a large number of NHR coregulators [16, 179], and differential expression between tissues could cause variation in NHR response. It is worthwhile to note,

however, that most NHR co-regulators appear to be rather general. The authors know of only one example of a tissue specific co-regulator; PGC-1 appears to exist exclusively in muscle, kidney, and liver cells [210]. Until further evidence of tissue specific co-regulators is identified, it must remain speculation that they are the chief explanation for selective response modifier effects. Moreover, coregulators do not act singly, but in large heteromeric collections that are bound to each other and to the NHR by one or a few "platform" coregulators. Cell-specific actions could be determined by the collective action of each heteromeric group. This in turn, would be driven by the presence and concentration of each coregulator [46]. The EAM shows how ligands could alter the choice or affinity for platform coregulators by positive or negative cooperativity within the NHR; thus, accounting for selective responses and even for cell/tissue-specific switching of agonist to antagonist.

Expression of different protein isoforms is a fourth possibility to explain different tissue activities of NHRs. Some NHRs have multiple isoforms of the intrinsically disordered N-terminus. Each N-terminal isoform may have a different intrinsic stability, which will result in a different sensitivity to coregulators [242], as shown for CBP and p300 interacting with GR [261]. Each isoform may also have different energies of interaction with the other domains. Simply expressing a different protein isoform could both change the sensitivity to NTD binding partners and the coupling response to allosteric regulators. Since NHRs

act as homo- or hetero-dimers, heteromers of NHR isoforms may alter the net response to a given ligand.

Post-translational modifications of NHRs include phosphorylation [94, 160, 318], sumoylation [232, 325], ubiquitylation [1, 424], and acetylation [112, 247]. Each modification has the potential to change NHR turnover, binding of ligands, and/or the coupling between NHR domains. For example, certain GR NTD phosphorylations—known to alter its transcriptional activity—also stabilize the NTD [435].

Considering the net effect of this long list of influences, the EAM may lead to an understanding of how the paradox of cell-specific selective response to a single steroid ligand occurs. EAM also explains how a given steroidal ligand can act as an agonist in one cell and an antagonist in another. A scheme for the application of EAM follows: First, the intrinsic coupling of a protein must be determined. Most of the exact values need to be known, but EAM can be used to estimate a few values if their sign and order of magnitude are known. Next, the model can be used to simulate the probabilities of the high affinity (active) states. These simulations can be done with and without ligands, and the effect of mutations can also be tested. A given mutation could affect the allostery between two domains (Δg of interaction) and/or the intrinsic stability of a domain (ΔG of the domain). Note that in these simulations the high affinity state is proportional to protein activity, or transcriptional activity in the case of NHRs. Lastly, after

simulating the results desired, one can test multiple mutations to find some that match the predicted outcome. Alternatively, if the interest is in drugs, then one can test multiple ligands for the desired coupling response.

1.6 Conclusions

The interplay between allostery and intrinsic disorder are just beginning to be unraveled for NHRs specifically, and transcription factors generally. It is clear thus far that these proteins use conformational dynamics to couple binding with allosteric effects. This potentially strengthens the cross talk between different receptor domains. The ensemble allosteric model is a theoretical framework that describes how a change in ensembles could drive allosteric effects, and it gives testable predictions for how NHRs should couple to their ligands and to themselves. Using the coupling response, the ensemble allosteric model allows one to predict the biological effect of drugs and coregulator proteins that bind NHRs. Positive regulators of transcriptional activity will have a positive coupling response with NHR domains that mediate transcriptional activity, and the opposite is predicted for negative regulators. This simple observation produces testable predictions for allosteric drug design and targeting. It also yields predictions that connect in vivo assays, measuring transcriptional activity, to in vitro assays, measuring protein stability.

Acknowledgments: VJH is supported by NIH grant GM063747 and NSF grant MCB-0446050. JTW was supported by the NIH training grant 5T32GM007231-39 to the Johns Hopkins University department of biology. HNM was supported by the NIH training grant T32-GM008403.

Chapter 2

Review of the Endosomal Sorting Complex Required for Transport (ESCRT)

2.1 Abstract

Endosomal maturation is a multi-step process that ends with the formation of a multivesicular body (MVB; see Figure 2.1). The internal vesicles of MVBs are laden with membrane proteins targeted for degradation, and formation of these internal vesicles was the first function ascribed to the Endosomal Sorting Complex Required for Transport (ESCRT). Disregulation of the ESCRT system can lead to cancerous growth and neuronal diseases, described herein [235, 244]. Furthermore, many budding viruses hijack ESCRT proteins in order to exit the cell [76, 140, 222, 277]. Understanding how the ESCRT system functions at a basic level is necessary to understanding ESCRT disease etiology and potential cures.

This chapter begins with a broad overview of endosomal membranes, the ESCRT system as a whole, and the function of ESCRTs in endosomal maturation. To that end, a large tableau has been made to help visualize endosomal maturation for the reader (Figure 2.3). Subsequent sections deal with the disease implications and the non-endosomal functions of the ESCRT proteins.

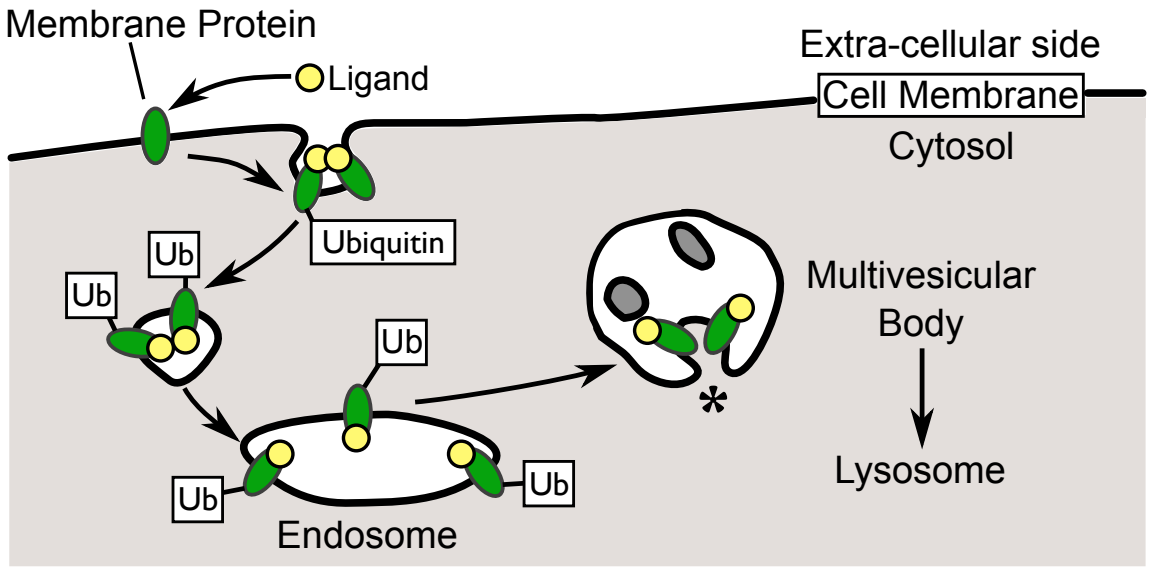


Figure 2.1 – An Overview of Endosomal Maturation
A membrane protein (green oval) binds its ligand (yellow circle) and is internalized (top left). The arrows show overall movement of protein/vesicles, **not** the exact membrane fission/fusion events. For membrane receptors, binding a ligand is often coupled with ubiquitination and internalization, see Figure 2.2 for details [180, 264]. Ubiquitin acts as a signal that recruits the ESCRT complex, and the critical ESCRT-mediated step is denoted by an asterisk. The multivesicular body (MVB) merges with a lysosome/vacuole to deliver proteins for degradation.

2.2 Endosomal Membranes

Endosomal maturation depends entirely on endosomal membranes, and endosomal membranes are contributed to by endocytosis—the process whereby cells invaginate part of their cell membrane to generate cytoplasmic vesicles. Endocytosis is a cellular mechanism to consume external materials and to downregulate, recycle, or degrade membrane receptors [271, 280]. There are both clathrin-dependent and independent forms of endocytosis [207, 271], and the two forms are regulated by over sixty proteins [280]. After endocytosis at the plasma membrane, there are numerous membrane fusion and fission events in which

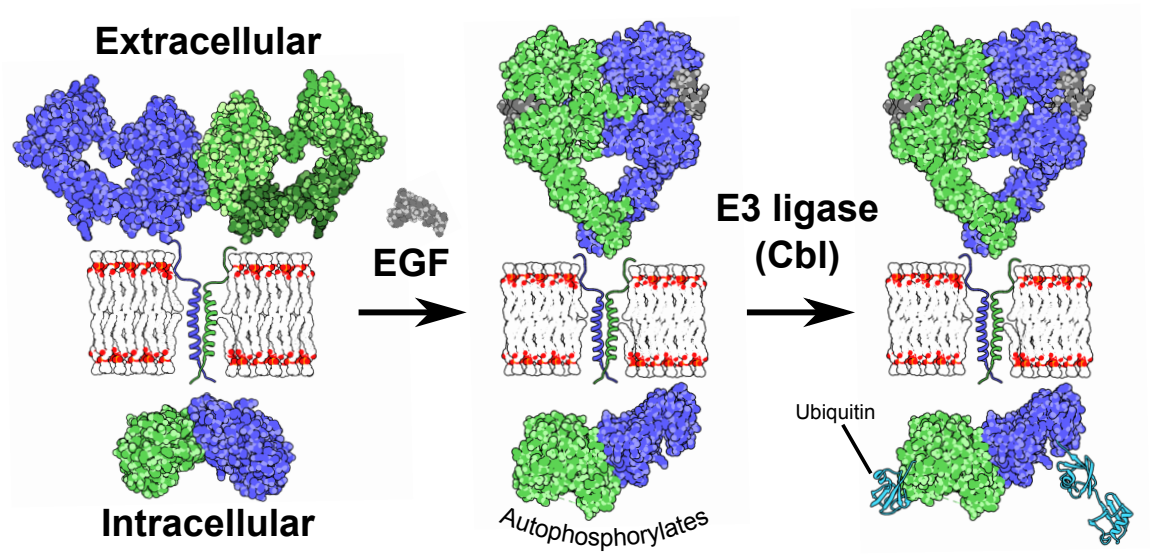


Figure 2.2 – **Activation of EGFR**
An inactive EGFR dimer (blue and green) is at the top. After binding EGF (gray), the whole protein goes through a large conformational change and the intracellular domains autophosphorylate. An E3-ubiquitin ligase, such as Cbl, eventually binds the phosphorylated intracellular domains and ubiquitinates EGFR [104, 162, 180, 264]. The phospholipid bilayer is depicted as an outline with white carbon, red oxygen, and orange phosphorous. The proteins are depicted with carbon as a darker color than the non-carbon atoms, except the transmembrane domains and ubiquitin (cyan) which are ribbons. Structures used here: [47, 105, 258, 382, 458]

cargo are sorted for either recycling back to the cell membrane or degradation in lysosomes. The author directs the interested reader to several reviews of these topics [181, 207, 271, 280]. For present consideration, it is the phospholipid composition of vesicles, particularly phosphatidylinositol phosphate (PIP), that defines endosomal membranes for the later recruitment of the ESCRT complex and other endosomal proteins [310].

The typical PIP modifications on the cell membrane are PI(4,5)P_2 and PI(3,4,5)P_3 [310]; whereas the early and late endosomes are defined by PI3P [125] and PI(3,5)P_2 [168, 355], respectively. It is via PIP's that much of the ESCRT machinery is targeted to endosomes; thus, it is important to understand how eukaryotic cells control phosphatidylinositol chemistry.

Initially, the PIP's on an internalized vesicle are converted to PI3P. A small GTPase protein called Rab5 recruits two phosphatases (4-Pase and 5-Pase) that remove 4- and 5-phosphates of inositol [371]. These phosphatases produce PI3P directly from PI(3,4,5)P_3 of the cell membrane and indirectly via PI(4,5)P_2 and a kinase that is also recruited by Rab5 (or dynamin-2 in phagocytosis [205]). The combination of the PI3 kinase, VPS34 [380], and the PIP phosphatases yields a concerted conversion of the early endosomal PIP pool to PI3P. (For information on the recruitment of Rab5: [7, 208, 209, 460]).

The PI3P of early endosomes binds to the FYVE Zn-finger domains of several ESCRT proteins and the PI5 kinase called PIKfyve [168, 239, 355] (FYVE is named

after some of the proteins with the domain: Fab1p, YOTB, Vac1, EEA1). PIKfyve's conversion of the late endosomal membrane's PI3P to PI(3,5)P_2 is important for the last stages of endosomal maturation. While the initial ESCRT complex binds PI3P, the last complex of the ESCRT system binds PI(3,5)P_2 ([439] and see ESCRT-III section below).[1]

2.3 Endosomal Maturation

The ESCRT proteins are broken up into roughly four complexes (ESCRT-0—III) [142]. There are also at least two additional proteins that are required for proper release of the ESCRT-III complex and the budding vesicle. All of these proteins are presented in tabular format in Table 2.1. For the rest of this chapter, the human nomenclature will be used. Please note that a few of the biochemical details differ between yeast and humans [169], but by-and-large the ESCRT system has been very highly conserved [241]. See Figure 2.3 for a structural overview of the ESCRT pathway.

[1] Because this thesis concerns nuclear actions of an otherwise cytosolic ESCRT protein, the author will briefly mention some of the nuclear actions of phosphoinositides (reviewed here: [34, 78, 137, 310]). Inositol-3-phosphate has been detected in nucleoli and VPS34 localizes to both the cytosol and nucleus [62, 125, 185]. VPS34 activity is required for EGFR transport to the nuclear membrane [85], but the exact function of VPS34 in the nucleus is still unclear. Some other PI kinases or their products have been shown to regulate chromatin, histone factors, and the transcriptional machinery [120, 387, 405]. Whether VPS34 is involved in any of these activities remains to be seen.

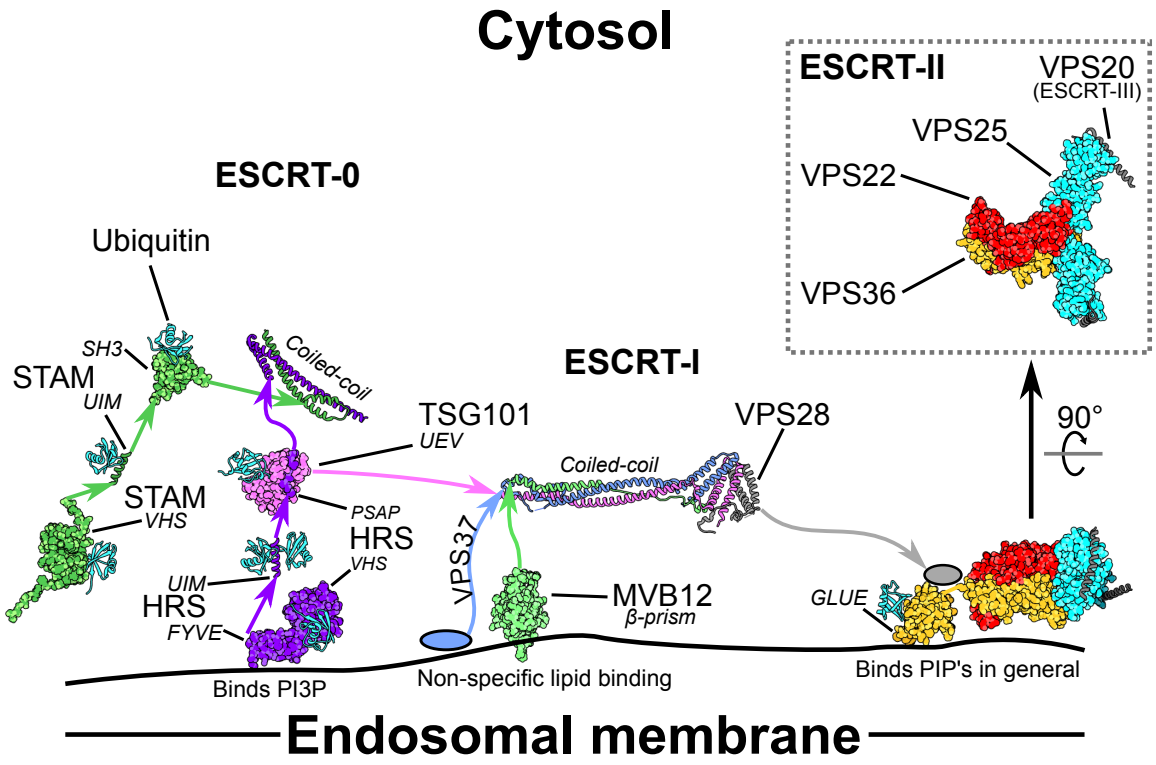


Figure 2.3 – **Cartoon Schematic of ESCRT-0—II**
Each complex is labeled in bold. Individual proteins are labeled with normal font and their subdomains are given a smaller, italicized font and sticks to point at the respective domain. Unstructured loops that connect domains are represented as arrows running from N- to C-termini and they are the same color as the respective proteins. Lipid binding specificities are described beneath lipid binding domains. Globular structures are depicted as atoms with carbon in a darker color. Helical structures and ubiquitins (cyan) are depicted as ribbons. Most ubiquitins are directly from the structures. The STAM-SH3 ubiquitin was roughly aligned with the NMR interaction surface, and the HRS-VHS ubiquitin is from aligning the STAM-VHS bound ubiquitin with HRS-VHS. Structures used here: pdb 3ZYQ and [6, 51, 152, 155, 169–171, 212, 223, 340, 393].

Complex	*S. cerevisiae* Protein Name	Metazoan Protein Name
ESCRT-0	Vps27	HRS (HGS)
	Hse1	STAM1, 2
ESCRT-I	Vps23 (Stp22)	**TSG101 (VPS23)**
	Vps28	VPS28
	Vps37 (Srn2)	VPS37A, B, C, D
	Mvb12	MVB12A, B or *UBAP1
ESCRT-II	Vps22 (Snf8)	VPS22 (EAP30)
	Vps25	VPS25 (EAP20)
	Vps36	VPS36 (EAP45)
ESCRT-III	Vps2 (Did4, Chm2)	VPS2A, B (CHMP2A, B)
	Vps20 (Chm6)	VPS20 (CHMP6)
	Vps24	VPS24 (CHMP3)
	Snf7 (Vps32)	SNF7A, B, C (CHMP4A, B, C)
	Vps60 (Chm5)	VPS60 (CHMP5)
	Did2 (Chm1, Vps46)	DID2A, B (CHMP1A, B)
	Ist1	IST1
	**Yj1049wp ?	CHMP7
Vps4-Vta1	Vps4	VPS4A, B (SKD1)
	Vta1	VTA1 (LIP5)
Bro1/ALIX	Bro1 (Vps31)	ALIX (AIP1) and HD-PTP (PTPN23)

Table 2.1 – **List of ESCRT proteins**
Commonly used names are listed first, with any alternate names in parentheses. Some of the metazoan proteins have duplicate genes that are listed by commas (A, B, etc.). This table is modified from the 2010 review by Hurley: [166]. TSG101 is highlighted because it is the focus of a subsequent chapter. *Even though UBAP1 is considerably different from the MVB12 genes, it can fulfill the MVB12 role in mammalian ESCRT-I [5]. **Yeast may or may not have a CHMP7 homologue. When CHMP7 was characterized, Yj1049wp was suggested to be a related protein in *S. cere.* [157].

2.3.1 ESCRT-0: Signal Recognition

The first ESCRT complex identifies ubiquitinated membrane proteins and corrals them onto the endosomal membrane, before passing the membrane proteins to the ESCRT-I complex. ESCRT-0 is comprised of just two proteins, HRS and either STAM1 or STAM2 (Hepatocyte growth factor-Regulated tyrosine kinase Substrate, Signal Transducing Adaptor Molecule) [142]. HRS tethers itself and STAM to the endosomal membrane, via its FYVE Zn-finger domain [381]. As stated above, FYVE domains preferentially bind the phospholipid, phosphatidylinositol 3-phosphate (PI3P) [239], which is enriched on the endosomal membrane [125].

In order to recognize ubiquitinated membrane proteins, the ESCRT-0 complex has multiple ubiquitin binding motifs. Both HRS and STAM have N-terminal VHS domains that bind one ubiquitin each (named after proteins with the domain: Vps27, HRS, STAM; [223, 339]). Both proteins also have ubiquitin interacting motifs (UIM), but the HRS version of UIM is able to bind two ubiquitins at once, versus the one ubiquitin that the STAM UIM binds [152, 284]. STAM also has a SH3 domain that is able to interact with one ubiquitin, but the ubiquitin is readily displaced by the deubiquitinating enzymes, UBPY and AMSH (Ubiquitin-Specific Protease Y and Associated Molecule with a Src Homolgy 3 domain; [155, 224]).

Deubiquitination by UBPY or AMSH early in endocytosis appears to promote recycling or delays degradation of targeted membrane proteins: both can deu-

biquitinate and promote recycling of EGFR [272, 283]. However, UBPY and AMSH are also necessary for efficient degradation of ubiquitinated membrane proteins [349, 394]. This second activity appears to be necessary for maintaining the pool of free ubiquitin [336] and is dependent on interactions with ESCRT-III proteins that will be explained in the subsection on the ESCRT-III complex.

Even though only one ubiquitin is necessary for a membrane protein to proceed through the ESCRT pathway [336], multiple ubiquitins likely assist the initial recognition of ESCRT cargo. Binding of ubiquitin to the ESCRT-0 complex is dependent partly on avidity (multiple binding events) and partly on the nature of ubiquitin linkages (different lysines) [223]. Individually, the VHS, UIM, and SH3 domains bind mono-ubiquitin with K_d's between 100–1000 µM [224, 339]. The ESCRT-0 complex as a whole binds mono-ubiquitin somewhat weakly, with a K_d of ∼920 µM [339]. In contrast, binding of polyubiquitin chains can have an apparent K_d closer to 20 µM [339], but this is highly dependent on the exact nature of the polyubiquitin chain.

Ubiquitin chains link using free amines and C-termini (i.e. amine from lysine of first ubiquitin→C-terminus of second ubiquitin). K63-linked ubiquitins have typically been associated with endosomal targeting of membrane proteins [229], but it is worth noting that some proteins can be targeted to lysosomes by either K63 or K48-linked ubiquitins [457]. Consistent with *in vivo* observations, the ESCRT-0 complex binds tetra-ubiquitin with the following linkage prefer-

ences [339]: K63 > K48 $\gg$N-terminus (apparent K_d = 18 µM, 43 µM, 140 µM respectively; Note: the author knows of no example where N-terminal linked ubiquitin causes ESCRT targeting, but the linkage is used in cell-signaling and proteasome targeting [342]).

Beyond the ubiquitin binding domains of both HRS and STAM are coiled-coil domains that form the ESCRT-0 heterodimer [340]; though, it has recently been reported that the lipid bound ESCRT-0 can form larger complexes on its own or with cargo [270, 398]. Lastly, HRS has a clathrin-binding domain on its C-terminus. HRS's ability to bind clathrin allows it to recruit clathrin coated-pits on endosomes [333], and this appears to help organize endocytosed receptors on the endosome [332].

2.3.2 ESCRT-I: Signal Transduction

The ESCRT-I complex acts to bridge the cargo-sensing ESCRT-0 with the ESCRT-II complex. ESCRT-I is assembled around one protein that acts as a central scaffold: Tumor Susceptibility Gene 101 (TSG101) [212]. TSG101 is recruited by HRS via the TSG101 ubiquitin E2 variant (UEV) domain [324]. The UEV domain is an inactive E2-ligase that is missing the reactive cysteine residue [23]. This allows the UEV domain to recognize ubiquitinated cargo without executing a ubiquitination reaction [200]. The UEV domain also contains a pocket for binding peptide motifs of the form "P T/S A P", with the second position being variable.

HRS contains a PSAP motif that is necessary for binding the TSG101 UEV domain, and this is the major connection between ESCRT-0 and ESCRT-I [324].

Following TSG101's UEV domain is a proline rich region that appears to be conformationally flexible and a coiled-coil domain that forms the core of the ESCRT-I complex [212]. To this core, two other proteins are recruited: MVB12 and VPS37 [212, 213]. Both MVB12 and VPS37 have non-specific membrane binding domains on their N-termini [51, 212], followed by coiled-coil domains that bind TSG101's coiled-coil, forming a heterotrimer. To the C-terminus of the ESCRT-I coiled-coil, is a cluster of α-helical hairpins formed by the C-termini of VPS37 and TSG101, and the N-terminus of VPS28 [212]. Lastly, the C-terminus of VPS28 binds the ESCRT-II complex and the first subunit of ESCRT-III, though the latter activity is not an absolute requirement for ESCRT-III recruitment [54, 73, 320].

2.3.3 ESCRT-II: Signal Amplification

The purpose of ESCRT-II is both signal transduction and amplification. The N-terminal GLUE domain (GRAM-like Ubiquitin binding in EAP45) of VPS36 and a basic helix on the N-terminus of VPS22 bind PIP's of the endosomal membrane, thereby localizing ESCRT-II to endosomes [6, 169, 374], regardless the presence of ESCRT-I [22]. The VPS36 GLUE domain also binds ubiquitin of targeted membrane proteins [6, 225], and immediately following the GLUE domain is a

helix that binds the C-terminus of the ESCRT-I complex component, VPS28 [169] (yeast maintain a slightly different mechanism [124]).

Following VPS36 and VPS22 are two VPS25 subunits that recruit ESCRT-III [22, 169, 225]. Note that all the previous ESCRT complexes bind 1:1 with each subsequent ESCRT complex. In contrast, for every ESCRT-II complex there will be two VPS20 subunits of ESCRT-III recruited to adjacent ends of the Y-shaped ESCRT-II structure [147, 169, 400] (Figure 2.3). Recruitment of two VPS20 proteins may simply increase the probability of recruiting one strand of ESCRT-III proteins or it may be necessary for formation of the native ESCRT-III complex [71, 369].

The binding of ESCRT-II to VPS20 also appears to prime the system for vesicle formation. Either ESCRT-II or VPS20 alone will bind to flat membrane surfaces, but the combination of ESCRT-II and VPS20 will preferentially bind membranes that curve away from the proteins [113]. Adding in the next ESCRT-III subunit, SNF7, even enables the complex to close open holes in a membrane monolayer [113]. Preferential binding of curved membranes may be due to conformational changes or wedging of the membrane by insertion of hydrophobic domains [67], but more recent work suggests that steric clashing of membrane-bound proteins is sufficient to explain membrane binding preferences [376].

2.3.4 ESCRT-III: Wrapping Things Up

The ESCRT machinery is part of what defines eukaryotic life [241], but one of the ESCRT complexes appears to be much more ancient than Eukarya. ESCRT-III orthologues mediate membrane scission during cytokinesis in some Archaea [92, 249] (a role conserved in Eukarya, as discussed later). Since the divergence of Archaea and Eukarya, the functions of ESCRT-III have expanded beyond cytokinesis. ESCRT-III executes the demand made by all the previous ESCRT complexes: that a vesicle, studded with membrane proteins marked for destruction, be cleaved off into the intralumenal space of the endosome.

Because all the ESCRT-III proteins have a similar domain architecture, it is easiest to first consider the properties they have in common: **1)** A core made of several helical-hairpins that arrange into either an “open” or a “closed” conformation depending on the context [32, 273, 296, 360], **2)** a positively charged surface on the core helices that binds phospholipid membranes [60, 296, 439], **3)** a hydrophobic N-terminal sequence involved in membrane binding [60, 453], **4)** a C-terminal motif that binds VPS4 [3, 139, 369], and **5)** a tendency to form filamentous homo- and hetero-polymers with one another [123, 143].

The field currently has only a incomplete picture of how membrane proteins are corralled into a tight, sterically unfavorable space on the endosomal membrane that is then exvaginated toward the lumen of the endosome by the

ESCRT-III complex (see Figure 2.4). The following description is a somewhat speculative reconstruction based on the most recent literature. A combination of ESCRT-0—II organizes cargo onto one locus [3, 52, 54, 320], then ESCRT-II binds two VPS20's which recruit two filaments of SNF7 to encircle the targeted membrane proteins [139, 143, 369, 400]. SNF7 polymerization continues and eventually incorporates VPS24 and VPS2 [21, 143, 228]. The combination of VPS24 and VPS2 causes the otherwise flat SNF7 spirals to deform into a three-dimensional projection that punches into membranes [71, 143]. At some point, the ESCRT-III proteins bind deubiquitination enzymes that remove ubiquitin from targeted membrane proteins [227, 349, 377] (ESCRT-0 can also recruit deubiquitination enzymes [155, 224]). The release of ubiquitin could achieve two goals: recycling ubiquitin and release of any bound ESCRT-0—I [336]. Eventually, VPS2 recruits VPS4 (next section) to constrict the ESCRT-III spirals further and release the whole ESCRT-III complex [3, 25, 228, 369].

The exact structure of ESCRT-III spirals is debatable. It is clear that the biologically relevant fibers involve an amalgamation of several ESCRT-III proteins that vary between the width of one or two ESCRT-III proteins [21, 71]. Most *in vitro* electron microscopy (EM) has used just one or two ESCRT-III's, often with the C-termini deleted and frequently in the absence of the membrane phospholipids that are integral to ESCRT-III function and structure [123, 228, 273]. Such studies often show tightly coiled rods of ESCRT-III proteins that are plausibly reminiscent

of structures seen *in vivo* in the last stages of vesicle formation [71, 139], and these protein aggregates are even reversible upon addition of VPS4 [123, 228]. It has been comparatively more difficult to understand the less constricted, early structures of ESCRT-III. A recent study using full-length SNF7 proteins assembled on membranes demonstrated that SNF7 spirals inwards because it lowers the free energy (ΔG) of the spiraling filament [369]. Without VPS4 present, the SNF7 C-termini sterically prevented the spirals from constricting and lowering their energy, but addition of catalytically inactive VPS4 (binds ESCRT-III C-termini) released that energy and produced tight coils of SNF7 [369].

2.3.5 VPS4 and VTA1: Resolution of Exvaginating Vesicles

It was shown about two decades ago that VPS4 is necessary for efficient sorting of membrane receptors to the yeast vacuole (mammalian lysosome equivalent) [24, 25]. VPS4 is a member of the AAA family of ATPases (ATPases Associated with diverse cellular Activities) and it forms hexameric assemblies [25, 391]. As long as its ATPase activity is present, VPS4 can de-polymerize ESCRT-III polymers and release nascent vesicles from the endosomal membrane [25, 71, 228] (Figure 2.4).

VPS4 has an N-terminal microtubule interacting and trafficking domain (MIT) that binds MIT interacting motif domains (MIM) of ESCRT-III [25, 83, 202]. It is primarily the MIM1 motif of VPS2 that is necessary for VPS4 recruitment, and the role of the other MIM's may be ancillary [3]. VPS4 is catalytically active *in*

vitro, but maximal activity requires a cofactor protein. Binding of VTA1 dimers around the VPS4 hexamer stimulates the ATPase activity of VPS4 [391, 446], and VTA1 also supports VPS4's binding of ESCRT-III via its own MIT domain that interacts with VPS60 and DID2 of the ESCRT-III complex [20].

2.3.6 ALIX

If Biology has a choice between doing something one way or two—it often errs towards redundancy. The ESCRT system is no different and the protein ALIX (Apoptosis Linked gene-2 Interacting protein X) is an alternative way for ESCRT-III to be recruited [73].[2] This alternative pathway was first described in viral budding [95, 388, 422], but it appears to be a natural part of ubiquitin-dependent and -independent endosomal maturation [73, 93, 198, 301, 311]. ALIX is similar to ESCRT-II in that it can bridge ESCRT-I and ESCRT-III [268], but ALIX does so in a completely different manner. In viral budding, ALIX binds ESCRT-I via the N-terminus of TSG101, and ALIX then recruits ESCRT-III by binding the C-terminus of SNF7 (ESCRT-II binds VPS28 and VPS20 instead) [197, 198, 274]. It has also been shown that ALIX can form dimers and that these dimers can bridge ESCRT-III filaments, forming ladder-like structures [321].

When ALIX was first discovered, it seemed as if ALIX was recruited by binding

[2] It is worth noting that mammals also express a second protein related to ALIX, called HD-PTP (His Domain Protein Tyrosine Phosphatase) [96]. HD-PTP seems to carry out functions similar to ALIX, and it has also been shown to bridge ESCRT-0 and ESCRT-III [8].

TSG101, and while this seems true in certain contexts (viral budding), the full story is much more complicated. ALIX can also bind ubiquitin, with a strong preference for the stereotypical K63 linkage seen on membrane proteins destined for the endosome [95]. The ability of ALIX to bind ubiquitinated membrane proteins is redundant with the ESCRT-0 complex, and to produce serious vacuolar defects, one must suppress both ALIX and ESCRT-0 [311]. Like ESCRT-0, ALIX can also bind phospholipid membranes; though its preference is for the endosomal lipid, lysobisphosphatidic acid (LBPA) [45]. With its combined ubiquitin, lipid, and ESCRT-III binding, ALIX can act as a redundant arm of the ESCRT pathway.

2.4 Non-Endosomal Functions of ESCRTs

Besides endosomal maturation, ESCRT proteins are also involved in cytokinesis [32, 65, 73, 234, 249], autophagy [107, 115, 235, 237, 348], cell and nuclear membrane repair [72, 89, 330], formation of exosomes [27, 347], formation of the cSMAC immune complex [419], chromatin condensation [385], transcription [74, 192, 293, 358, 370, 431], and mRNA localization [173].

ESCRT involvement in cytokinesis proceeds in a manner very similar to endosomal maturation, except TSG101 and ALIX are recruited by CEP55 to the cytokinetic midbody [234]. The end result is recruitment of ESCRT-III to form the cortical filaments that constrict during abscission [133]. ESCRT activity in

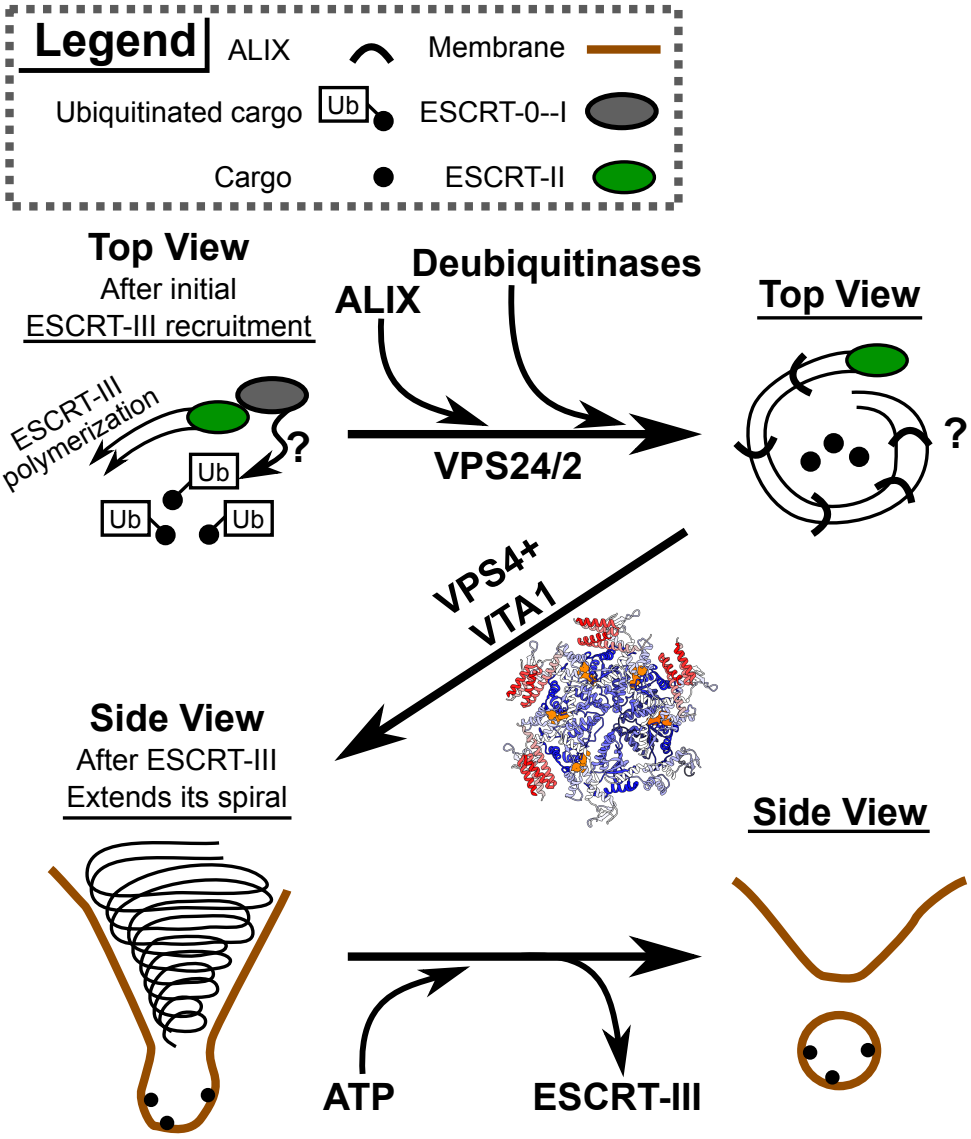


Figure 2.4 – **Speculative Cartoon of ESCRT-III Vesicle Formation**
In the top left, ubiquitinated membrane proteins are corralled by uncertain mechanisms that likely depend on ESCRT binding their ubiquitins. Initial ESCRT-III polymerization by SNF7 likely cordons off the targeted proteins. Recruitment of later ESCRT-III proteins (VPS24/2), ALIX, and deubiquitinases may lead to a situation like the top right image. ALIX forms ladders between ESCRT-III filaments *in vitro*, but the *in vivo* significance and organization are unknown [321]. VPS24/2 promote three-dimensional spiralling [143], and VPS2 recruits VPS4 that also constricts ESCRT-III spirals [369]—the transition from the top right to bottom left is one possible rendition. In the last step, VPS4's ATPase activity disassembles ESCRT-III, releasing the monomers to solution [123, 228]. The VPS4+VTA1 structure is shown as a ribbon cartoon in the center, with VPS4 as blue and white, VTA1 as red, and ATP as orange. See also: [71, 139, 369, 391].

autophagy and membrane repair appears to proceed in a similar manner, just with different initial recruitment steps that are still under investigation.

Exosomes and microvesicles are extracellular vesicles that are often filled with proteins or RNA for the purpose of cell-to-cell signaling ([334, 418]). The control of exosome formation is still poorly understood, but this is an exciting field with much potential for cancer diagnostics because some cancers use exosomes to promote tumor growth and metastasis [373]. Extra-cellular vesicles can form directly on the cell membrane (microvesicles) or via fusion of an MVB with the cell membrane (exosomes) [334]. While both of these processes sound exactly like ESCRT activity, some exosomes require ESCRT [27, 347], while others do not [414].

ESCRT-II has some unexpected side-roles in metazoan cells. VPS36 binds *bicoid* mRNA during metazoan development and is necessary for proper localization of the mRNA [173]. The mammalian ESCRT-II can also bind a transcription elongation factor (ELL) and promote its activity *in vitro* [192, 358], but after almost twenty years no further work has been done and many details remain unknown. For example, it is unknown whether VPS36 binds nascent mRNA during transcription or why the entire ESCRT-II complex is involved. In a related problem, TSG101 has been implicated in the transcriptional activity of steroid hormone receptors; yet, it is unknown why. All current experiments have used viral promoters or artificial transcriptional assays [74, 153, 293, 431]. It is un-

known what genes are controlled naturally by TSG101, nor is it known whether TSG101 recruits other ESCRT proteins after binding to transcription factors. It is tempting to think that TSG101 could indirectly recruit ESCRT-II for transcription elongation, but no such interaction has been shown.

Besides ESCRT proteins having moonlighting functions, there are a large number of endocytic proteins that also have secondary functions in the nucleus or elsewhere (reviewed here: [319, 328]).

2.5 Disease Implications

Inhibition or mutation of the ESCRT system can lead to membrane protein mislocalization [25, 343] and even cancerous growth [244, 285]. The cancerous phenotype seen in metazoans is likely due to upregulation of growth factor receptors in the absence of the ESCRT system [23]. Disregulation of ESCRT or its binding partners has also been associated with neuronal diseases, likely due to autophagy or cytokinetic defects (frontotemporal dementia [235], hereditary spastic paraplegias [352]).

As mentioned earlier, numerous budding viruses are known to hijack ESCRT proteins in order to exit the cell membrane (in a microvesicle or exosome-like manner). These viruses include some of the greatest scourges known to humankind: HIV, Ebola, Hepatitis B and C and E, Herpes, Rabies, Rous sarcoma

virus, and many others [76, 140, 222, 248, 277, 297, 388]. In an interesting twist, Epstein-Barr virus even uses TSG101 to activate its genes and ESCRT to exit the *nucleus*, as opposed to the cell membrane [74, 233]. As terrible as these viruses can be, early studies that used viruses as models of ESCRT action revealed a great deal of what is now known about ESCRT. There have been attempts to develop anti-virals based on ESCRT interactions [204], but the author knows of no ESCRT-based anti-viral—nor would it be easy to develop a therapy without side-effects because viruses have hijacked the very interactions necessary for natural ESCRT function [324].

2.6 Conclusion

ESCRT proteins are a defining feature of eukaryotic life [241]. It is the ESCRT complex that gives eukaryotes exquisite control over intracellular membranes, thereby allowing me to segregate reactions or signals to/from neighboring cells or give negative feedback to those same signals. With improved control over cell-to-cell communication, eukaryotes have produced the most obvious examples of multi-cellular life. It took millions of years to evolve, and yet, in just the past thirty years, much has been discovered about the ESCRT complex. The author looks forward to seeing what the next thirty years reveal.

Chapter 3

Structural Stability of the coiled-coil domain of Tumor Susceptibility Gene (TSG)-101

This chapter is largely a quote of material published in Biochemistry with Dmitri Toptygin, Randy Cohen, Natalie Murphy, and Vincent J Hilser. Dr. Toptygin did the Time Correlated Single Photon Counting experiments and related analyses detailed here. Mr. Cohen and Ms. Murphy assisted

in protein purification. Prof. Hilser provided overall guidance, including the design of experiments and analyses. See [438]

3.1 Abstract

The Tumor Susceptibility Gene-101 coiled coil domain (TSG101cc) is an integral component of the endosomal maturation machinery and cytokinesis, and also interacts with several transcription factors. The TSG101cc has been crystallized as a homotetramer but is known to interact with two of its binding partners as a heterotrimer. To investigate this apparent discrepancy, I examined the solution thermodynamics of the TSG101cc. Here, I use circular dichroism, differential scanning calorimetry, analytical ultracentrifugation, fluorescence, and structural thermodynamic analysis to investigate the structural stability and the unfolding of the TSG101cc. I demonstrate that TSG101cc exists in solution primarily as a tetramer, which unfolds in a two-state manner. Surprisingly, no homodimeric or homotrimeric species were detected. Structural thermodynamic analysis of the homotetrameric structure and comparison with known oligomeric coiled-coils suggests that the TSG101cc homotetramer is comparatively unstable

on a per residue basis. Furthermore, the homotrimeric coiled-coil is predicted to be much less stable than the functional heterotrimeric coiled-coil in the Endosomal Sorting Complex Required for Transport-I (ESCRT-I). These results support a model whereby the tetramer–monomer equilibrium of TSG101 serves as the cellular reservoir of TSG101, which is effectively outcompeted when its binding partners are present and the hetero-ternary complex can form.

3.2 Introduction

Tumor Susceptibility Gene-101 (TSG101, yeast homologue vps23) is a protein with multiple roles in eukaryotic biology. As a member of the Endosomal Sorting Complex Required for Transport-I (ESCRT-I) complex, TSG101 recruits other proteins to maturing endosomes and assists in the formation of multivesicular bodies [23, 343, 346] and many budding viruses [288]. TSG101 is also involved in cytokinesis [102], and in mammals, TSG101 can act as a transcriptional regulator [432]. The numerous roles of TSG101 are facilitated through its four domains (Figure 3.1): The N-terminal UEV domain recruits TSG101 to membrane proteins targeted for degradation [200], the proline-rich region recruits TSG101 to the cytokinetic furrow [102], the coiled-coil assembles most of the ESCRT-I complex [212], and the steadiness (i.e., stability) box plays a role in forming the bridge between the ESCRT-I and ESCRT-II complexes [124, 212]. Of these,

the coiled-coil domain is also of significance in transcriptional regulation, as it is believed to interact with several transcription factors [293, 432]. In spite of the growing body of literature identifying novel binding partners and previously unknown functions of TSG101, there is an incomplete picture regarding the structural thermodynamics of the TSG101 coiled-coil and its role in modulating the stoichiometry of interaction.

The coiled-coil of TSG101 (TSG101cc) has been crystallized as a homotetramer (PDB 3iv1), but its oligomerization state has yet to be established in solution. This issue is significant because contrary to the tetrameric form observed for TSG101cc in isolation, the solution and crystal structures of the ESCRT-I coiled-coil are heterotrimers of TSG101/vps23 (yeast homologue) and two binding partners [212, 287]. Note that other coiled-coils have been shown to adopt different structures or oligomeric states in crystal and in solution form [129, 451]. Furthermore, ongoing studies in the Hilser lab suggest TSG101cc can also form a heterodimer, with the human glucocorticoid receptor (to be published elsewhere). Biological regulation through the modulation of the oligomeric state of a protein has been reported in a number of other systems (morpheein proteins [176]), including oligomerization of helix-loop-helix and coiled-coil peptides, interactions in the mothers-against-decapentaplegic family of proteins, the hepatitis B viral capsid protein, and porphobilinogen synthase among others [15, 57, 66, 103, 177, 389]. Here, I set out to determine the oligomeric state of TSG101cc in solution

and to determine whether its oligomeric state is functionally modulated.

I show that TSG101cc in solution exists primarily as a tetramer that is in equilibrium with unfolded monomers. Despite the ability of TSG101cc to form heterotrimeric coiled-coils, no measurable homotrimeric (or homodimeric) species were observed. Thermal unfolding of TSG101cc as a function of protein concentration and pH provided access to the thermodynamic mechanism of stabilization of the tetramer. Furthermore, structural thermodynamic analyses of the TSG101cc tetramer and ESCRT-I coiled-coil heterotrimer illuminates the structural basis of the tetramer stability, and provides insight into the role of the homotetramer in regulating its hetero-oligomeric function.

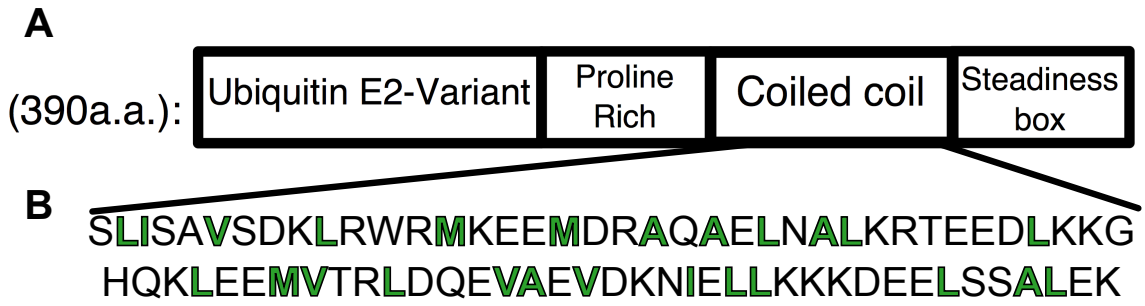


Figure 3.1 – **Schematic of TSG101 Domain Organization**
A) The domains pictured here are scaled to their length and the domain of interest is expanded in **B)** The coiled-coil construct used in this study comprises amino acids 229-304 of the native sequence (76 amino acids long plus 2 a.a. left by the N-terminal tag). Hydrophobic residues are colored green.

3.3 Methods

3.3.1 Protein Expression and Purification

The sequence of my expression construct spans amino acids 229-304 of TSG101 (Figure 3.1, appendix, UniProt: Q99816). On the N-terminus is a 9xHis tag, followed by a TEV protease cleavage site and an extra serine. Competent BL21 DE3 pLysS cells were made in house and used for expression (Novagen strain). Growth proceeded in LB medium, with a 1 mM IPTG induction at an O.D.600 0.6-1.2, followed by growth at 37 ºC for four hours or overnight at 16 ºC. Little difference in expression was observed under these conditions. Cells were pelleted, washed with PBS, and lysed by sonication (20 mM Tris, 6M Gdn, 20 mM imidazole, 500 mM NaCl, pH 8). After clearing the lysate at 15 krpm for an hour, the lysate was purified using Ni-NTA. The His tag was cleaved off using TEV protease, and protein was passed over an anion exchange column followed by Ni-NTA. At this point TSG101cc was a single band by SDS-PAGE [68]. For differential scanning calorimetry (DSC), protein was purified further using size exclusion chromatography (SEC, HiLoad 16/600 Superdex 75 pg, GE). An example SDS-PAGE of SEC purified TSG101cc is shown in Figure 3.2.

For pyrene labeled protein, used in analytical ultracentrifugation and fluorescence (supplement), I purchased PMIA (N-(1-pyrenemethyl) iodoacetamide;

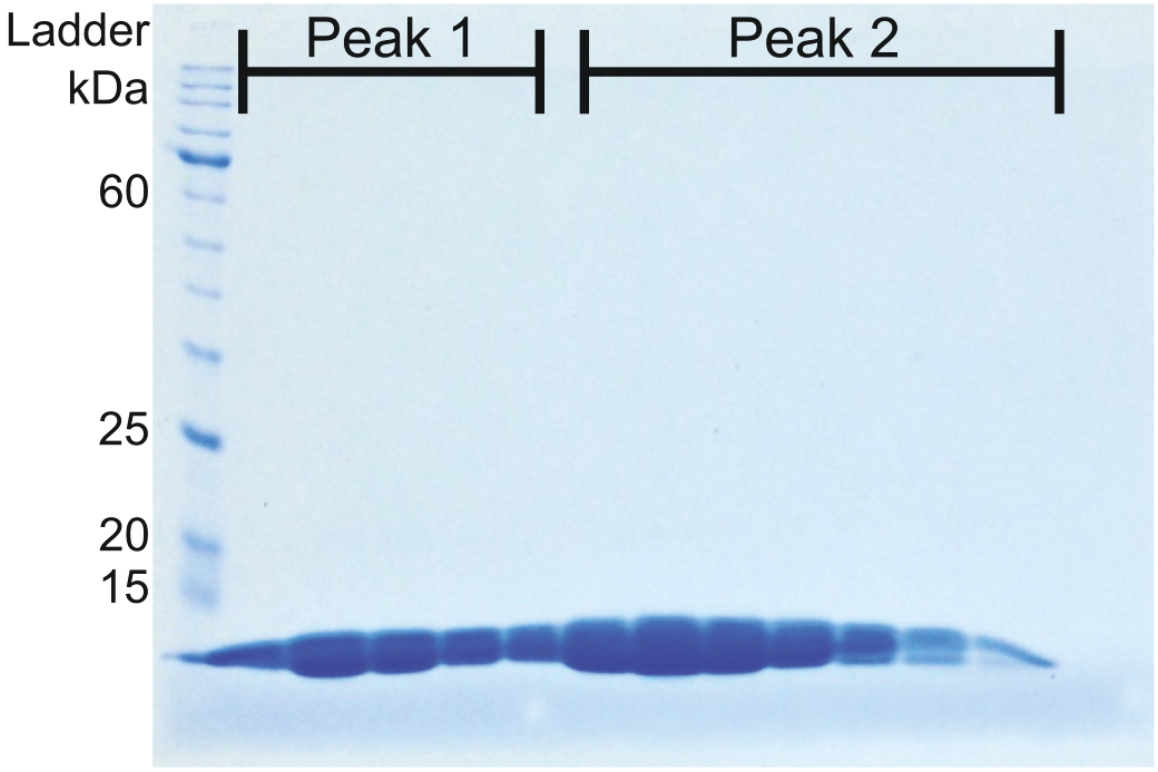


Figure 3.2 – **SDS-PAGE gel of SEC purified TSG101cc**
The protein shown here is unlabeled (lacks an engineered cysteine) and was used for DSC. Expected MW is 9 kDa. Two peaks were often seen in SEC of unlabeled protein, supporting the two-state hypothesis for TSG101cc.

Setareh biotech). Quikchange PCR was used to change a serine to a cysteine in my expression construct (supplement). The serine in question is part of the N-terminal tag, not the native TSG101 sequence. This protein was purified as described above, dialyzed to labeling buffer (50 mM HEPES and 1 mM TCEP at pH 7.3), then labeled for one hour at room temperature using 10 molar equivalents of PMIA, relative to moles of protein, dissolved in DMSO. For efficient labeling, it was found that 5% DMSO was required in the final reaction mixture. Labeling was quenched using 10 molar equivalents of DTT, relative to the moles of PMIA. This resulted in a turbid solution that was 0.2 µm filtered, then separated using SEC.

The first peak of SEC was consistently a mixture of TSG101cc with one or as many as three pyrene labels, as determined by mass spectrometry (supplement). Iodoacetamide is known to react with amines and attempts to prevent this by pH optimization did not succeed. However, the second peak I observed from SEC was consistently a mixture of singly labeled (66-80%) and unlabeled protein. This is what was used in all analytical centrifugation and fluorescence experiments described here.

The concentration of unlabeled protein was determined by its absorbance at 280 nm in guanidine, $\varepsilon_{TSG101cc}$ = 5,690 $M^{-1}cm^{-1}$ [100]. The concentration of PMIA labeled protein was determined by its absorbance at 345 nm (44,700 $M^{-1}cm^{-1}$ Setareh Biotech). For PMIA labeled protein, extinction coefficients at

other wavelengths were set relative to the 345 nm pyrene peak.

3.3.2 Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI)

All MALDI was done using a Bruker AutoFlex III MALDI-Time of Flight/ToF. The identity of purified protein was verified by trypsin digest (Promega sequencing grade) followed by reflectron mass spectrometry. Labeling efficiency of TSG101cc with pyrene was determined by linear Time of Flight (ToF) mode. The labeling position was verified by trypsin digest with a combination of reflectron mode and "LIFT" mode [390]. Data were analysed by mMass and then curated by hand.

3.3.3 CD

For CD measurements, protein was dialyzed to 20 mM Na_2HPO_4 plus 50 mM NaCl and pH adjusted with HCl (6.7, 7.2, 7.5). All CD data shown here were gathered using an Aviv CD spectrophotometer and a 1 mm path length, quartz cuvette. Each wavelength scan was in 1 nm increments, with a bandwidth of 1 nm, from 250 to 195 nm. The thermal melt data shown are with 1 ºC increments and a 2 minute incubation at each step. All CD data were averaged for five seconds at each measurement and buffer signal was subtracted. Reversibility (often between 70—96%) was determined by remeasuring the signal at 222

nm and 20 ºC. I obtained my best data by diluting the protein several hours beforehand and degassing the protein for 5 minutes before measurement.

3.3.4 DSC

Protein was dialyzed to 20 mM PIPES and 200 mM NaCl and pH adjusted with NaOH (6.7, 7.2, 7.5). Data were collected with a microCal DSC at a scan rate of 1.5 ºC per minute and duplicated with independent protein preparations and at least two scans. Repeated scanning of TSG101cc was highly reversible (99% refolded) as long as the final temperature was ≤ 73 ºC (346 K). The data shown here have been scan rate normalized, buffer subtracted, and then normalized to the total concentration of monomeric protein. To circumvent baseline uncertainty in the native region that rendered determination of ΔC_p problematic, the ΔC_p was held constant at a value of 690 cal per K*mol of monomer based on the COREX analysis.

3.3.5 Fluorescence

PMIA-labeled protein was dialyzed to the same buffer used in AUC and the concentration was determined in the same manner. All fluorescence experiments were done at 20 ºC.

Time-resolved fluorescence measurements were made using picosecond laser excitation with a Time-Correlated Single Photon Counting (TCSPC) setup; the

instrument is described elsewhere [412]. The exceptionally long excited-state lifetime of pyrene compelled me to slow down the exciting pulse rate from 4.1 MHz to 820 kHz and to extend the time window of the TCSPC electronics from 25 ns to 100 ns. Time-Resolved Anisotropy (TRA) data and Time-Resolved Emission Spectra (TRES) were collected using this instrument. The excitation wavelength was close to 310 nm for both types of experiments; exciting light was vertically polarized.

In TRA experiments, vertically and horizontally polarized fluorescence was measured at an emission wavelength of 377 nm, where a sharp emission peak of pyrene monomer is located; the broad emission band of pyrene excimer appears at longer wavelengths. The TRA data analysis was done using a previously described program polartcp [413].

In TRES experiments fluorescence emission was measured through a magic-angle polarizer; the emission wavelength was stepped from 367 to 517 nm at 5 nm increments. TRES of a system undergoing a chemical reaction was first described by Laws and Brand [231]. In the case of the PMIA-labeled protein, excited-state pyrene dimers (excimers) are formed on the time scale of fluorescence decay, which is revealed by the TRES. The data analysis using several programs described elsewhere [412] generated the amplitude spectra, which represent the wavelength variation of the amplitudes associated with different exponential terms in fluorescence intensity decay.

Steady state fluorescence was measured using a SLM-Aminco SLM-48000S spectrofluorometer with polarizers at the magic angle. Excitation was at 329 nm (8 nm monochromator bandwidth) and emission was gathered from 345—550 nm (4 nm bandwidth). A Corning glass filter type 0-54 was used in series with the exciting monochromater to cut out all tryptophan-exciting wavelengths that were present due to off-band monochromater transmission. After non-specific adsorption of PMIA-labelled protein to quartz was discovered at sub-μM concentrations, I coated my cuvette first with BSA (data that is shown), to little effect, then with silane. Even after silanization, I still observed a significant loss of protein at the lowest concentrations. This adsorption was only observed in the sub-μM data of Figure S2, and was not percieved in any of the other experiments of this manuscript.

3.3.6 AUC

PMIA-labeled protein was dialyzed to 20 mM HEPES, 200 mM NaCl, and 1 mM TCEP and adjusted to a pH of 7.2 using NaOH. Data were collected at initial concentrations of 5, 9, and 18.2 μM pyrene-labeled TSG101cc, and the samples were loaded into epoxy centerpieces with sapphire windows (Spin Analytical). The Beckman XL-I was set to take absorbance measurements at 329, 336, and 345 nm. The three rotor speeds used were 29, 37, and 44 krpm, and final equilibrium took about 18 hours per rotor speed. The data were fit globally with several

models using HeteroAnalysis (JL Cole and JW Lary of University of Connecticut Bioservices Center). Partial specific volume and solvent density were calculated using SEDNTERP (J Philo). The data were replicated using a separate protein preparation that yielded fitted values within error of those presented here.

3.3.7 COREX/BEST Calculations

Per-residue stability analysis of TSG101cc proceeded in a manner largely described before for other proteins [148, 307, 440]. In this work, I used a window size of eight residues and 100,000 Monte Carlo simulations per partition. The entropy weighting factor was set to 1.029 to make the maximal residue stability roughly match my CD data at 150 µM protein and a pH of 7.2. I used H++ to calculate pK_a's for titratable residues, and used this as input for a pH dependent COREX/BEST calculation [14]. Solvent accessible surface area calculations were made with a 1.4 Å rolling ball from a subroutine of COREX [290], and subsequent thermodynamic analyses proceeded similarly to the manner described by others before [50, 294, 409].

3.3.8 DSC and CD Fitting Equations

Fitting of DSC and CD requires determination of the fraction folded ($\mathcal{F}_{\mathcal{U}}$) at every temperature [110, 184], and can be referenced to moles of oligomer [246]

or moles of monomer as described below. The basic relationships are:

$$\mathcal{N}_n \overset{\mathcal{K}}{\rightleftharpoons} n * \mathcal{U}$$

$$\mathcal{K} = \frac{[\mathcal{U}]^n}{[\mathcal{N}_n]}$$

$$\text{Total protein concentration} = \mathcal{P}_{\mathcal{T}} = n * [\mathcal{N}_n] + [\mathcal{U}]$$

$$\text{Fraction of protein in state U} = \mathcal{F}_{\mathcal{U}} = \frac{[\mathcal{U}]}{[\mathcal{U}] + n * [\mathcal{N}_n]}$$

The quantity $\mathcal{F}_{\mathcal{U}}$ can be solved by substituting into the equilibrium constant solved for $[\mathcal{U}]$: $[\mathcal{U}] = [\mathcal{N}_n]^{1/n} * \mathcal{K}^{1/n}$, noting that $\mathcal{K} = e^{-\Delta \mathcal{G}/\mathcal{R}*\mathcal{T}}$ is represented in cooperative units of oligomer. Recasting the equilibrium expression in terms of moles of monomer yields: $k = \mathcal{K}^{1/n} = e^{-(\Delta \mathcal{G}/n)/\mathcal{R}*\mathcal{T}}$, where $\Delta g = \Delta \mathcal{G}/n$ and corresponds to the Gibbs free energy per monomeric mole of protein.

The equilibrium can thus be represented as:

$$k = \frac{[\mathcal{U}]}{[\mathcal{N}_n]^{1/n}} \tag{3.1}$$

Given $\mathcal{F}_{\mathcal{U}}$ denotes the fraction of monomeric protein and $(1 - \mathcal{F}_{\mathcal{U}})$ denotes the oligomeric fraction of protein:

$$[\mathcal{U}] = \mathcal{P}_{\mathcal{T}} * \mathcal{F}_{\mathcal{U}} \tag{3.2}$$

$$[\mathcal{N}_n] = (1 - \mathcal{F}_\mathcal{U}) * \mathcal{P}_\mathcal{T}/n \tag{3.3}$$

Substituting $[\mathcal{U}]$ from Equation (3.2) and $[\mathcal{N}_n]$ from Equation (3.3) into Equation (3.1) yields:

$$\frac{\mathcal{F}_\mathcal{U}}{(1 - \mathcal{F}_\mathcal{U})^{1/n}} = \frac{\mathcal{P}_\mathcal{T}^{\frac{1}{n}-1} * k}{n^{1/n}} \tag{3.4}$$

The roots of Equation (3.4) can be solved for $\mathcal{F}_\mathcal{U}$.

Before enumerating the fitting equations derived from Equation (3.4), I define the reference concentration as the concentration where $\Delta S_{trans} = 0$ such that the term $\frac{\mathcal{P}_\mathcal{T}^{\frac{1}{n}-1}}{n^{(1/n)}} = 1$, with concentration having no effect on k in Equation (3.4). In thermodynamic terms, ΔS_{trans} is the gain in entropy due to the increase in translational rotational degrees of freedom upon dissociation.

From the right side of Equation (3.4), one can derive an explicit entropy term by rearranging $\mathcal{P}_\mathcal{T}$ and n to be part of the exponential that k represents $e^{-\frac{\Delta g}{\mathcal{R}*\mathcal{T}}}$. Converting the right side of Equation (3.4) to oligomeric units [183]:

$$\left(\frac{\mathcal{P}_\mathcal{T}^{\frac{1}{n}-1} * k}{n^{1/n}}\right)^n = \frac{\mathcal{P}_\mathcal{T}^{1-n} * \mathcal{K}}{n} = \frac{\mathcal{K}}{n * \mathcal{P}_\mathcal{T}^{n-1}} \tag{3.5}$$

Substituting for K (i.e., $\mathcal{K} = e^{-\Delta \mathcal{G}/\mathcal{R}*\mathcal{T}}$) in Equation (3.5), and expressing in exponential form:

$$e^{-\Delta \mathcal{G}/\mathcal{R}*\mathcal{T}} * e^{ln[(n*\mathcal{P}_\mathcal{T}^{n-1})^{-1}]}$$

And from above: $e^{-\Delta \mathcal{G}/\mathcal{R}*\mathcal{T}} * e^{-\mathcal{R}\mathcal{T}\, ln[(n*\mathcal{P}_\mathcal{T}^{n-1})^{-1}]/\mathcal{R}\mathcal{T}}$

From above: $e^{-\{\Delta G + \mathcal{R}T\,ln[(n*P_T^{n-1})^{-1}]\}/\mathcal{R}*T}$

From above, given $\Delta G = \Delta H - T\Delta S$: $e^{-\{\Delta H - T\Delta S + \mathcal{R}T\,ln[(n*P_T^{n-1})^{-1}]\}/\mathcal{R}*T}$

From above, note that $\mathcal{R}$ is in the same units as entropy:

$$e^{-\{\Delta H - T(\Delta S - \mathcal{R}\,ln[(n*P_T^{n-1})^{-1}])\}/\mathcal{R}*T}$$

$$\text{Thus: } \Delta S_{trans} = -\mathcal{R} * ln[(n * P_T^{n-1})^{-1}] \tag{3.6}$$

Note that for $n * P_T^{n-1} = 1 \rightarrow \Delta S_{trans} = 0$ ($P_T = 0.25^{1/3}$ Molar for tetramer), and the concentration where translational entropy is zero will be the same regardless of whether one uses a monomeric or oligomeric reference. Above that concentration, ΔS_{trans} will weakly favor binding and below that concentration the converse is true. Because Equation (3.4) takes these terms into account, all the fitted Δg values are independent of concentration and equivalent to Δg at the hypothetical concentration where ΔS_{trans} is zero.

For DSC, I fit the average excess change in heat capacity, which is the sum of equations (3.7-3.9) below, namely the base and transition excess heat capacities:

$$Base\ C_{p,N} = (1 - F_U) * a \tag{3.7}$$

$$Base\ C_{p,U} = F_U * b \tag{3.8}$$

where a and b are empirical folded and unfolded baseline constants. My native baselines were not good enough to directly fit the ΔC_p. The transition heat capacity is described by the derivative of F_U with respect to temperature:

$$Transition\ C_p = \frac{\delta F_U}{\delta T} * \Delta H \tag{3.9}$$

An analytical solution to $\frac{\delta F_U}{\delta T}$ can be solved using Equation (3.4), above. My derivative differs in form from the derivative described by Professor Freire in 1989, but the two equations are equivalent and their computational efficiency is nearly the same.

$$\frac{\delta F_U}{\delta T} = \frac{n^{1-1/n} * P_T^{\frac{1}{n}-1}(1 - F_U)^{\frac{1}{n}+1}}{n(1 - F_U) + F_U} * \frac{\Delta H}{R * T^2} * e^{-\delta g/RT} \tag{3.10}$$

For CD, I fit the following equation in a global manner, using F_U derived from the roots of Equation (3.4), above (for n = 4):

$$Average\ signal\ (T) = (m_N * T + c_N) * (1 - F_U) + (m_U * T + c_U) * F_U \tag{3.11}$$

The m and c parameters are empirical fits to the native and unfolded baselines.

3.3.9 Fluorescence Methods

PMIA-labeled protein was dialyzed to the same buffer used in AUC and the concentration was determined in the same manner. All fluorescence experiments were done at 20 ºC.

Time-resolved fluorescence measurements were made using picosecond laser excitation with a Time-Correlated Single Photon Counting (TCSPC) setup; the instrument is described elsewhere [412]. The exceptionally long excited-state lifetime of pyrene compelled me to slow down the exciting pulse rate from 4.1 MHz to 820 kHz and to extend the time window of the TCSPC electronics from 25 ns to 100 ns. Time-Resolved Anisotropy (TRA) data and Time-Resolved Emission Spectra (TRES) were collected using this instrument. The excitation wavelength was close to 310 nm for both types of experiments; exciting light was vertically polarized.

In TRA experiments, vertically and horizontally polarized fluorescence was measured at an emission wavelength of 377 nm, where a sharp emission peak of pyrene monomer is located; the broad emission band of pyrene excimer appears at longer wavelengths. The TRA data analysis was done using a previously described program polartcp [413].

In TRES experiments fluorescence emission was measured through a magic-angle polarizer; the emission wavelength was stepped from 367 to 517 nm at 5 nm

increments. TRES of a system undergoing a chemical reaction was first described by Laws and Brand [231]. In the case of the PMIA-labeled protein, excited-state pyrene dimers (excimers) are formed on the time scale of fluorescence decay, which is revealed by the TRES. The data analysis using several programs described elsewhere [412] generated the amplitude spectra, which represent the wavelength variation of the amplitudes associated with different exponential terms in fluorescence intensity decay.

Steady state fluorescence was measured using a SLM-Aminco SLM-48000S spectrofluorometer with polarizers at the magic angle. Excitation was at 329 nm (8 nm monochromator bandwidth) and emission was gathered from 345—550 nm (4 nm bandwidth). A Corning glass filter type 0-54 was used in series with the exciting monochromater to cut out all tryptophan-exciting wavelengths that were present due to off-band monochromater transmission. After non-specific adsorption of protein to quartz was discovered, I coated my cuvette first with BSA, to little effect, then with silane. Even after silanization, I still observed a significant loss of protein at the lowest concentrations.

3.4 Results

3.4.1 MALDI mass spectrometry

To confirm the identity of my protein, matrix assisted laser desorption ionization (MALDI) mass spectrometry was carried out. The general experimental set up involves firing a UV laser at a protein sample that is embedded in an acidic and UV absorbing matrix (α-Cyano-4-hydroxycinnamic acid is common for small MW samples). The positively charged sample is then ejected from the sample plate, which has a high and positively-charged voltage placed on it. The velocity of the sample is obtained from its time of flight (ToF) in the MALDI sample chamber, before striking the MALDI detector. The number of charges on a given analyte can be discerned from its isotopic distribution, but is usually assumed to be +1 for MALDI. The mass of the sample is obtained thusly: $\mathcal{V} * q = \mathcal{E} = 0.5 * \mathcal{v}^2 * m$, where $\mathcal{V}$ is the voltage applied to the MALDI plate, q is charge (assumed +1 in this case), $\mathcal{E}$ is energy, $\mathcal{v}$ is velocity, and m is mass.

There are several modes of MALDI detection: **1)** Linear mode uses ToF alone to determine the mass of large molecules (MW >5,000). **2)** Reflectron mode uses an electrical field at the top of the ToF path to reflect ions back down and at a different detector. This has a selective effect and greatly improves resolution of small MW analytes, but the paths of large MW analytes are not bent as much

by the electric field and thus cannot be analyzed because they hit the wall of the chamber. **3)** Laser Induced Fragmentation (LIFT) uses electrical fields to select a particular ion that is then fragmented using a very high UV laser power [390]. Breaking up large molecules into their constituent parts helps identify the parent molecule.

MALDI of trypsin digested TSG101cc often gave a sequence coverage exceeding 90%, see part A of Figure 4.5. Identification of pyrene labeled TSG101cc proceeded in three steps: Linear MALDI-ToF of intact protein, reflectron MALDI-ToF of trypsin digested protein, and LIFT MS/MS on the N-terminal fragment identified by reflectron MALDI. These data are presented in parts B–D of Figure 4.5. The data show that size exclusion chromatography can separate mono-labeled protein from multi-labeled protein. Further MS on the mono-labeled protein showed that the pyrene moiety is attached to one of the three N-terminal amino acids of TSG101cc, which includes the intended cysteine target.

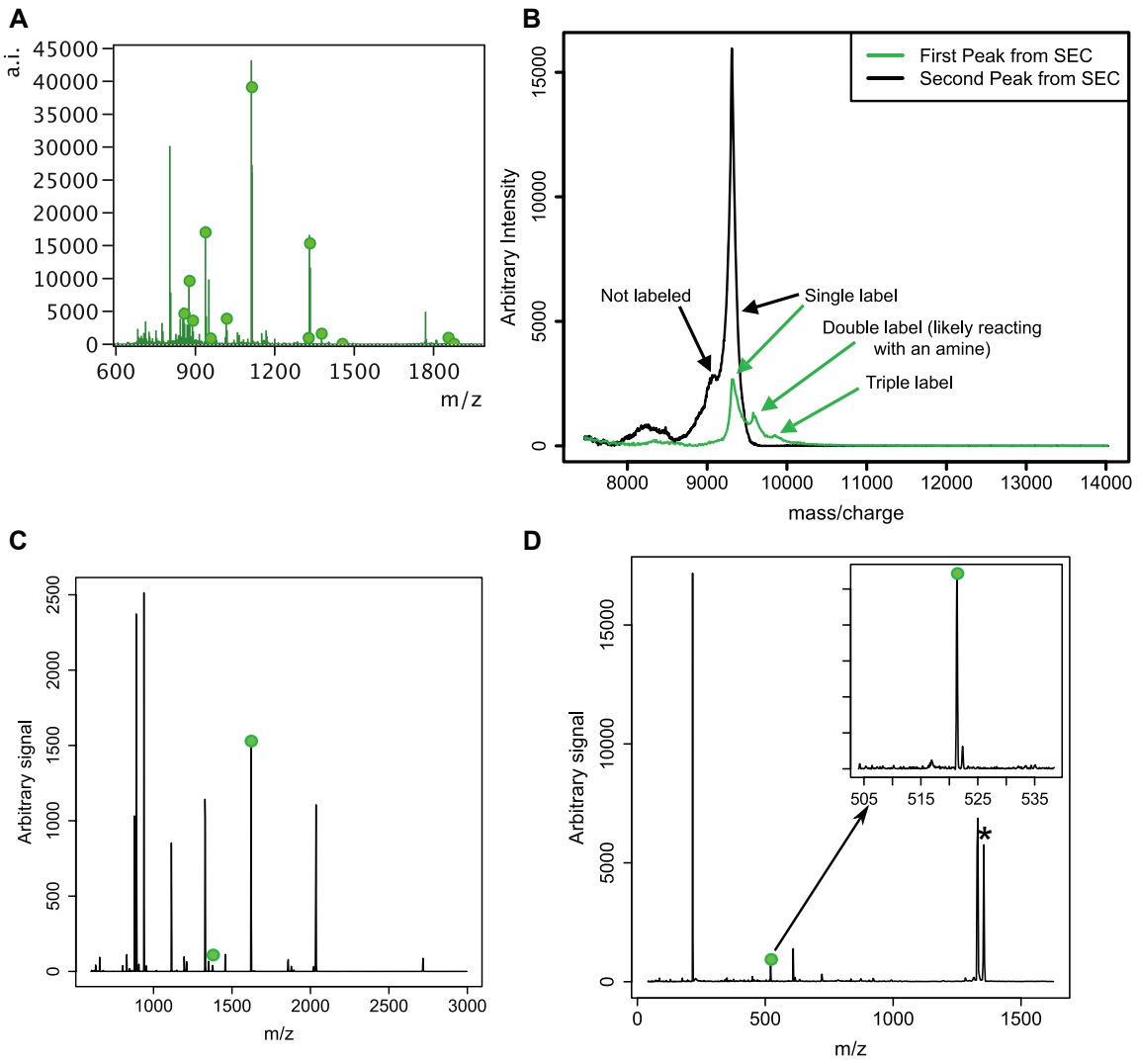


Figure 3.3 – MALDI Data Verifying Identity of TSG101cc Protein Used in This Study
All data are represented as arbitrary units of intensity on y-axis and mass/charge units on the x-axis. **A)** Trypsin digested TSG101cc without pyrene routinely provided over 90% coverage. Peaks consistent with TSG101cc are labeled with green circles. **B)** Labeling of TSG101cc with pyrene led to two peaks on a size chromatography column. The first peak off the column, colored green, is a mixture of differently labeled species. The second peak, colored black is mostly singly labeled protein. **C)** Trypsin digestion of pyrene labeled protein yielded two fragments consistent with pyrene on the N-terminus of TSG101cc (green circles). Note that the cysteine was engineered into the N-terminal tag. **D)** LIFT of the fragment 1350 m/z units produces one fragment consistent with the first three amino acids (GCS) and pyrene acetamide. Parent ion is labeled with an asterisk, daughter ion is labeled with a green circle. Inset: A zoomed view of the ion of interest.

3.4.2 Circular Dichroism (CD) Reveals α-Helical Structure

NOTE: all CD and DSC were with unlabeled TSG101cc (no pyrene and no cysteine engineered into the N-terminus)

Circular Dichroism (CD) was used to investigate the secondary structure of TSG101cc. As Figure 3.4 reveals, and consistent with the crystal structure, TSG101cc is indeed highly α-helical at low temperatures, as indicated by the strong peaks in negative ellipticity at 208 and 222 nm. Heating of the TSG101cc from 4 - 68 ºC (Figure 3.4 resulted in a thermally-induced conformational transition in which the α-helical content was lost, giving way to a CD spectrum that is essentially invariant above 56 ºC and which corresponds to the expected spectrum of a disordered polypeptide (Figure 3.4). Analysis of the temperature-dependent transition reveals an isodichroic point at 204 nm, which suggests that the thermal transition, to a first approximation, involves

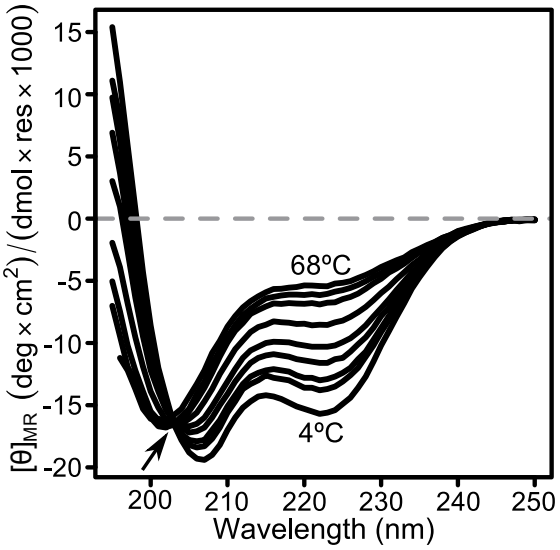


Figure 3.4 – Circular Dichroism Wavelength Scan of a Temperature Melt of TSG101cc
The data have an isodichroic point at 204 nm, highlighted by the arrow. This particular data set is from 24 μM protein that refolded to 72% of its original signal at 20 ºC. The start (4 ºC) and end (68 ºC) point temperatures are labeled. The intermediate temperatures after 4º: 14º, 20º, 26º, 32º, 38º, 44º, 50º, 68 ºC. Two temperatures, 56º and 62º, are omitted for clarity because they overlapped with the 68º scan.

just two states. Consistent with this observation, further analysis of the temperature dependence of the CD spectra using singular value decomposition (SVD) [172, 384] reveals that just two singular values can explain 97.7% of the variation in the data (see Figure 3.5 for a replotting of the first two singular values).

To determine whether changes in the degree of oligomerization are associated with the conformational transition, thermal unfolding experiments were performed at various protein concentrations [184, 404]. The data were fit globally with respect to temperature, concentration, and pH using the roots of Equation (3.4). As Figure 3.6 reveals, both protein concentration and pH affect the temperature dependence of the mean residue ellipticity ($[\theta]_{MR,\ 222nm}$). The increase in apparent transition temperature (T_m) with concentration reveals that unfolding decreases the degree of oligomerization. The shift of the apparent T_m (for a given protein concentration) to lower temperatures with increasing pH, suggests that the α-helical oligomeric state binds protons

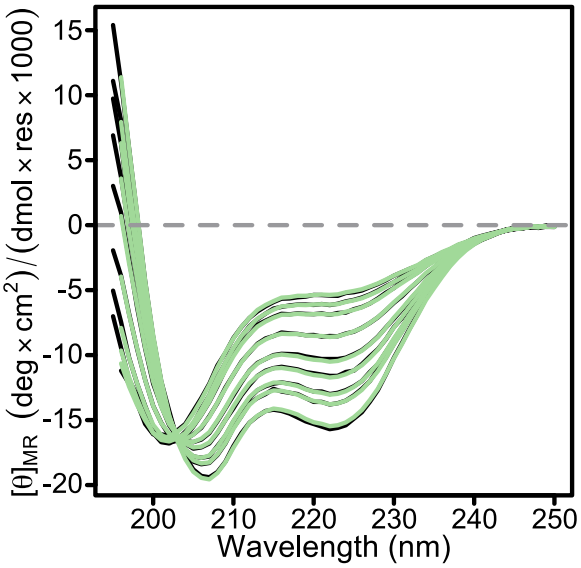


Figure 3.5 – Singular Value Decomposition of Circular Dichroism Melt Data as a Function of Wavelenth
The original data are plotted as black lines underlaying the singular value decomposition, in light green lines (same line thickness). The original data extend to a lower wavelength for context. Only the first two singular values were used for this figure: 222.4 and 59.1 with a total trace value of 288.1

more tightly than the unfolded state. Specifically, the average number of protons lost upon unfolding the protein can be determined using the expression [326]:

$$\Delta n_{protons} = \frac{\delta \mathcal{T}_{ref}}{\delta p\mathcal{H}} * \frac{\Delta \mathcal{H}}{2.303 * \mathcal{R} * (\mathcal{T}_{ref})^2} \tag{3.12}$$

where Δn is the change in the number of protons bound and $\frac{\delta \mathcal{T}_{ref}}{\delta p\mathcal{H}}$ is the slope of a plot of the reference temperature, $\mathcal{T}_{ref}$, versus pH. The data indicate a loss of $\approx$0.6 protons by each monomer upon unfolding, which corresponds to $\sim$2 protons per tetramer. The global fit of the data according to a two-state (folded tetramer to unfolded monomer) transition produced the parameters shown in Table 3.1.

pH	T_{ref} Kelvin	ΔH kcal/mol	ΔCp, cal/(mol*K)	Average $\Delta n_{protons}$
6.7	350.4±3.4	56.5 ± 9.8	675 ± 421	-0.64
7.2	347.1±3.4			-0.65
7.5	345.4±3.4			-0.66

Table 3.1 – Fitted Thermodynamic Values for CD of TSG101cc
Five concentrations from 25–150 μM were used at each pH. Globally fitted values are presented with 95% confidence intervals of the fit and are with respect to moles-monomer. The change in protons bound was calculated with Equation (3.12). Note that ΔC_p was determined from the temperature dependence of ΔH obtained at the different pH values. The fitted ΔH° is the intrinsic enthalpy of unfolding where $\Delta S_{translation}$ = 0, and is thus shared for all pH values.

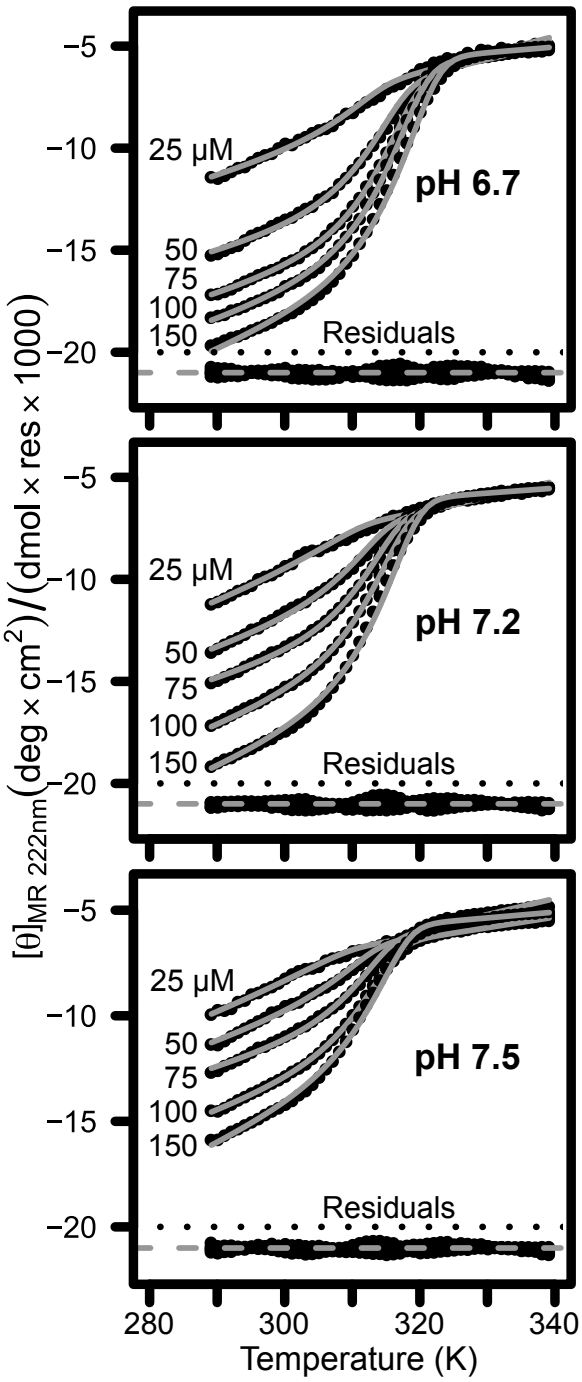


Figure 3.6 – **Circular Dichroism Thermal Melt at Multiple pH's and Concentrations**
μM concentrations and pH are marked alongside each dataset. The best fits to a two-state, tetramer unfolding transition are plotted over the data points and the residuals are plotted below each dataset. Refolding for 10 minutes at 20 ºC returned 70-96% of the original signal.

3.4.3 Differential Scanning Calorimetry (DSC) Comports with CD Data

To further analyze the thermodynamics associated with the thermally-induced unfolding-dissociation transition, DSC was performed on TSG101cc as a function of concentration and pH. Because DSC directly monitors the heat associated with the thermally-induced transition, it is unique among experimental approaches in that it provides direct access to the partition function (i.e. the significantly populated states and their respective energies). Specifically, the unfolding and dissociation of an oligomeric protein $\mathcal{N}_n$ will be reflected in the shape of a DSC curve. Increasing n will cause an increase in the asymmetry and sharpness of a DSC curve [110], and increasing the number of states will cause broadening and/or extra peaks [111].

Consistent with the results obtained from CD, analysis of the DSC thermograms reveals that TSG101cc exists as a tetramer that undergoes a cooperative unfolding/dissociation reaction with temperature Figure 3.7. Also similar to the CD experiments, lower concentrations and higher pH values were found to destabilize the TSG101cc tetramer. Because the T_m concentration dependence is logarithmic, the concentration-induced T_m shifts in Figure 3.7 are subtler than those seen at the lower concentrations used in CD. Shown in Table 3.2 are the average values from individual fits of the DSC data, using the same equations

derived for the CD analysis. Importantly, the DSC data are in general agreement with the CD data.

pH	T° Kelvin	ΔH° kcal/mol	Average Δnprotons calculated
6.7 (4)	358.9±5.6	45.6 ± 1.6	-0.29
7.2 (4)	356.7±8.4		-0.30
7.5 (2)	355.9±2.2		-0.30

Table 3.2 – Fitted Thermodynamic Values for DSC of TSG101cc
Data were fit individually and averages of the fits are presented ± two standard deviations. Enthalpy is with respect to moles-monomer. Parentheses indicate the number of independent experiments used in the fits and averaging. ΔC_p was held at a constant 690.085 (estimated by COREX). The fitted ΔH° is the intrinsic enthalpy of unfolding where $\Delta S_{translation} = 0$. See text for further description. Proton binding was calculated with Equation (3.12).

3.4.4 Fluorescence Indicates that Pyrene Labeling of TSG101cc Does not Disturb its Structure Significantly

I labeled the TSG101cc with pyrene in order to use fluorescence techniques to investigate its dynamics and dissociation at low concentrations. If two or more pyrene labels are in close proximity in the ground state, then once a pyrene label is excited, it can form a dimer with another pyrene label in the ground state, creating an excited state dimer, known as an excimer. The known structure of the TSG101cc tetramer has two N-termini next to one another on either end of the coiled-coil; thus, I expected pyrene excimers to form if labeling the protein had negligible effect on its structure.

I carried out three experiments: steady state fluorescence, time-resolved

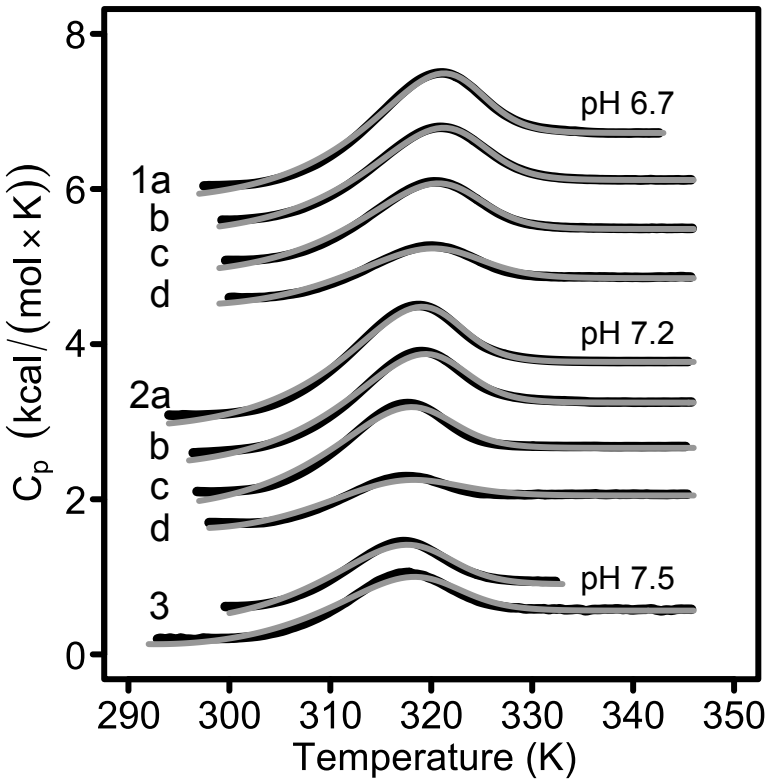


Figure 3.7 – Differential Scanning Calorimetry of TSG101cc at Various pH's and Concentrations
All of these data are normalized to the moles of monomeric protein and offset vertically for visual purposes. Data are black lines and the associated fits are grey lines. Concentration and pH are as follows: (1) pH 6.7 a) and b) 414 μM, c) 295 μM, d) 200 μM; (2) pH 7.2, a) 590 μM, b) 414 μM, c) 295 μM, d) 147 μM; (3) pH 7.5, both at 416 μM. Each scan was reproduced with nearly 99% of the protein refolding.

anisotropy (TRA), and time-resolved emission spectra (TRES). The goal of the first experiment was initially to determine the equilibrium binding constant, but the TSG101cc tenaciously adsorbs to quartz at exceedingly low concentrations, even after multiple attempts to prevent this. Nonetheless, a dissociation event occurs at μM concentrations, as the broad pyrene excimer emission band disappears by 50 nM (Figure 3.8). This is qualitatively consistent with the AUC data (see next section), where the fraction of tetrameric protein is estimated be <1% at

100 nM.

TRA data were collected to investigate the overall structural dynamics of the labeled TSG101cc. Data were fit to the following equations:

$$\mathcal{I}_v(t) = g^{+\frac{1}{2}}[1 + 2r(t)]\mathcal{I}(t) \tag{3.13}$$

$$\mathcal{I}_h(t) = g^{-\frac{1}{2}}[1 - r(t)]\mathcal{I}(t) \tag{3.14}$$

$$r(t) = \sum_{m=1}^{\mathcal{M}} \beta_m * e^{-t/\phi_m} \tag{3.15}$$

$$\mathcal{I}(t) = \sum_{n=1}^{\mathcal{N}} \alpha_n * e^{-t/\tau_n} \tag{3.16}$$

Where $\mathcal{I}_v$ and $\mathcal{I}_h$ are the experimentally observed vertical and horizontal fluorescence intensities, respectively, and g equals the ratio of the instrument sensitivities to vertical and horizontal polarization. Unlike reference [413], where g played the role of a free fitting parameter, in this work the value of g was measured independently and fixed. $\mathcal{I}(t)$ and $r(t)$ are independent of one another and are the total fluorescence intensity and the time-resolved anisotropy, respectively. β is the zero time anisotropy for each correlation time, ϕ. α is a weighting parameter for each fluorescence lifetime, τ.

Our data demonstrate that the pyrene label is attached at a single, homogenous location, as it has only on correlation time (0.022 in Table 3.3). The protein's rotation is described by three correlation times that suggest a highly extended

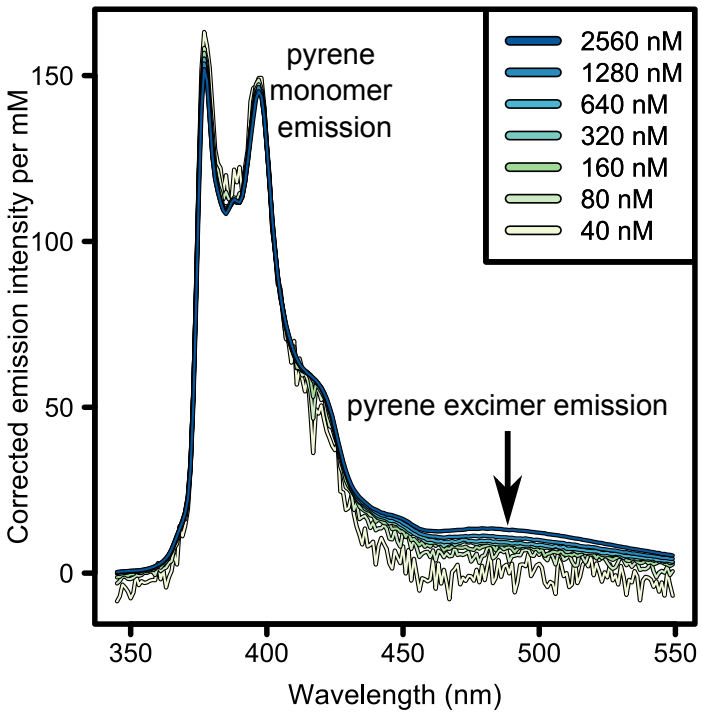


Figure 3.8 – **Steady State Fluorescence of TSG101cc-pyrene**
As TSG101cc-pyrene is diluted, the pyrene excimer fluorescence disappears (≈500 nm). This is indicative of an oligomerization reaction and is consistent with the crystal structure.

shape; however, my data does not perfectly match the theory of Tirado for a perfect cylinder's rotation [408]. My TRA data are consistent with a cylinder of between 9.1 Å and 13.2 Å in diameter, and between 152.4 Å and 142.3 Å in length. The crystal of the tetramer has a diameter of 30 Å and a length of 134 Å; given the concentration range I was able to use (5—20 μM), this is most consistent with a mixture of tetrameric and monomeric protein (consistent with AUC data). See Figure 3.9 for example polarized TCSPC data curves, Table 3.3 for fitted correlation time parameters, and Table 3.4 for the fitted intensity decay parameters of the TRA model equations (3.13–3.16).

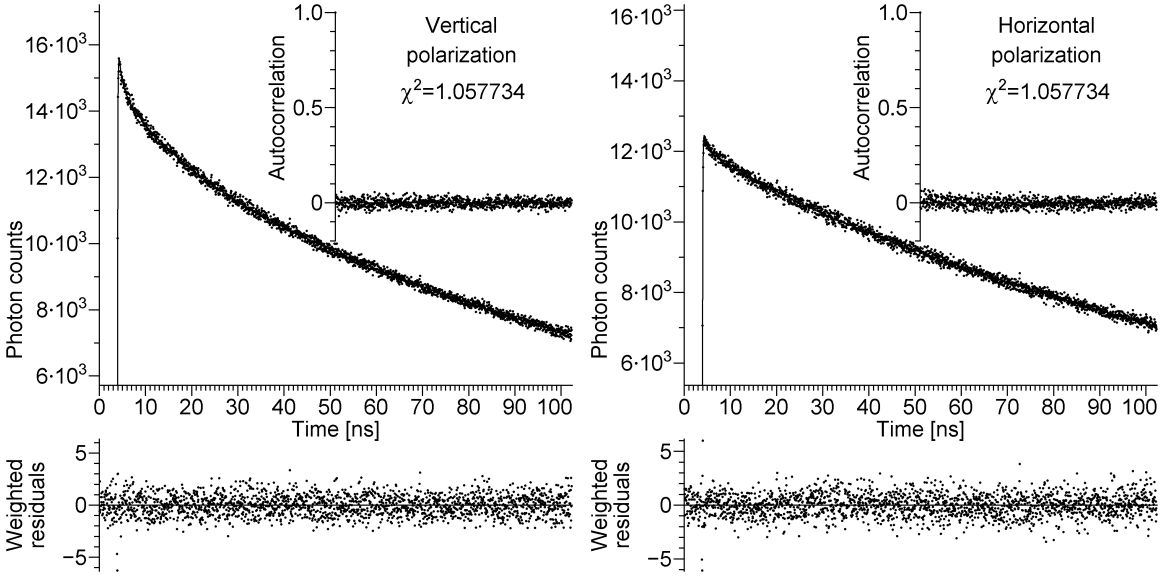


Figure 3.9 – **Polarized TCSPC Data for TSG101cc-pyrene**
This particular dataset is from protein at 20 μM and the associated fit can be seen in Table 3.3 and Table 3.4. Data are dots and fits are black lines. The figure on the left is emission intensity for horizontally polarized light as a function of time and the figure on the right is the same except it is for vertical emission. The best fit to this set of data is four exponentials describing anisotropy and three describing lifetime.

Our TCSPC instrument registered photons within a 100 ns time window that

Table 3.3 – Fitted Correlation Time Values

ϕ (ns) ± standard deviation	β ± standard deviation
0.0220 ± 0.0346	0.309747 ± 1.053504
2.4176 ± 0.7006	0.017678 ± 0.003157
15.8632 ± 7.1479	0.018337 ± 0.004394
69.1623 ± 8.0822	0.045040 ± 0.006541

Table 3.4 – Fitted Intensity Decay Values

τ (ns)	% Intensity	% α
0.5683	0.00028	0.07806
14.2614	0.00684	0.07653
187.4985	0.99289	0.84540

was too short as compared to the excited-state lifetime of pyrene, which is of the order of 300 ns. A longer time window would result in a significant loss of time resolution, which would make it impossible to measure the rotational correlation time for the pyrene label. Also it was technically impossible to expand the time window beyond 100 ns [412]. With the 100 ns time window it was impossible to resolve the decay of the pyrene monomer (300 ns) from that of the pyrene excimer (80 ns) when the emission at just one wavelength was analyzed. Thus, the 187.5 ns lifetime in Table 3.4 represents two unresolved exponentials corresponding to the monomer and the excimer. To overcome this and glean better information on excimeric pyrene, I collected time-resolved emission spectra that were then

fit by the following equation:

$$\mathcal{F}(\lambda, t) = \sum_{n=1}^{\mathcal{N}_{exp}} \alpha_{\lambda n}(\lambda) * e^{-t/\tau_n} \tag{3.17}$$

Where $\mathcal{N}_{exp}$ exponential terms possess the same lifetimes, τ_n, at each wave-

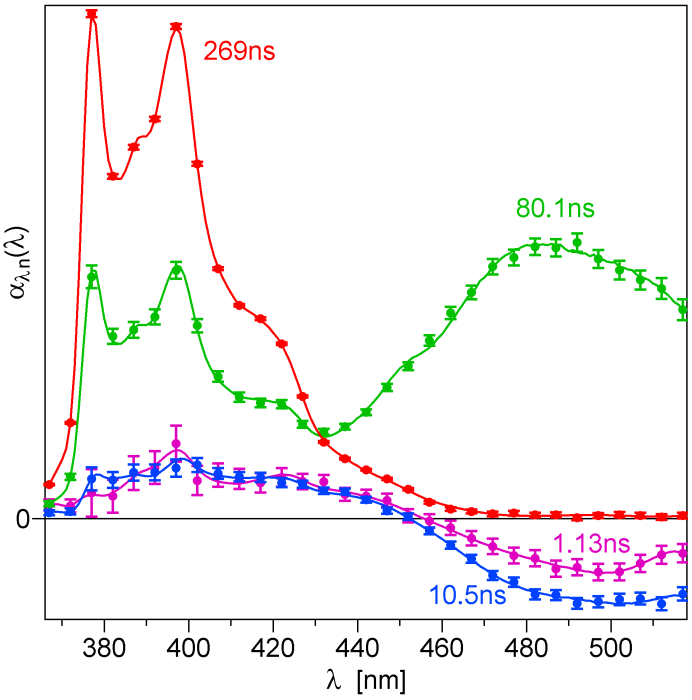


Figure 3.10 – Decay-Associated Amplitude Spectra of TSG101cc-pyrene
The TCSPC data were fit by the model in Equation (3.17). The decay time τ_n is printed near each curve using the same color. Amplitudes α_n calculated directly from TCSPC data are shown as points with error bars and the fit obtained using program trspectr [412] is shown as lines. The red line is monomeric pyrene population, unable to form excimer. The green line is excimer forming pyrene population. The dynamics of equilibration between monomer and excimer are described by the two exponentials depicted by the blue and purple lines. The negative amplitudes seen in the blue and purple lines describe the build up of excimer emission during the first few nanoseconds after the end of the excitation pulse.

length, λ, but the corresponding amplitudes α_n vary with wavelength and are therefore called amplitude spectra, $\alpha_n(\lambda)$. My data are described by a series of four exponentials Figure 3.10: One describes a population of monomeric

pyrene that is unable to form excimers (τ = 269 ns), this includes pyrene labels attached to proteins in a monomeric form and pyrene labels attached to proteins that oligomerized with an unlabeled counterpart. A second describes a population of pyrene that is able to form excimer (τ = 80.1 ns); these pyrene labels are attached to oligomerized proteins with labeled counterparts. The third and fourth exponentials describe the equilibration between monomeric and excimeric pyrene (τ = 1.13 ns and 10.5 ns). Pyrene excimers do not exist immediately after the exciting pulse, but appear several nanoseconds later as a result of an excited-state reaction. Since the distance between pyrene labels at the time of excitation is not uniform, the process of excimer formation is multiexponential (I see two exponentials given my time resolution). The negative amplitudes at the wavelengths of excimer emission describe the increase in the excimer emission at early times, when the excimers are formed. These data confirm the long lifetime of pyrene, and more importantly, demonstrate TSG101cc oligomerization in a manner consistent with the crystal structure. The fluorescence data suggest that pyrene does not perturb the structure of TSG101cc to a significant extent. I was also able to observe dissociation of the tetramer, but adsorption prevented attempts to extract an accurate equilibrium constant in Figure 3.8. Note that adsorption was only perceivable for the lowest concentrations in Figure 3.8, and had no apparent effect on the other data presented here.

3.4.5 Analytical Ultracentrifugation (AUC) Confirms the Two TSG101cc States

Equilibrium AUC was also used to probe TSG101cc's oligomeric state. Thermal denaturation has the potential to conceal unstable intermediates, as the unfolded state rapidly increases in probability with increasing temperatures. Centrifugation of TSG101cc at 20 ºC, however, allows me to challenge the findings obtained from CD and DSC. To increase the sensitivity to partially unfolded or different oligomeric states, TSG101cc was end-labeled with pyrene using a Cys residue incorporated into the N-terminal tag (see appendix for sequences), and detection was facilitated with absorbance optics.

Shown in Figure 3.11 is a representative AUC trace at 9 µM. A two-state fit of the data reveal a monomer-tetramer equilibrium that is consistent with the CD and DSC results: MW = 9778±295 Daltons, stoichiometry = 3.97±0.11, $\ln(K_d)$ = -39.41±1.51, K_d =7.66×10-18 M^3, RMSD = 0.01003. Of note is that a one-state fit is inconsistent with the known monomeric mass (9313 Da monomeric mass by MALDI, see Figure 4.5) and produces skewed residuals at higher rotor speeds (MW = 35775±102 Daltons, RMSD = 0.01101). Similarly, a three-state fit was precluded as it produced no justifiable improvement of the residuals over the two-state fit. Importantly, the K_d, as measured by AUC corresponds to a stability of 23.0 ± 0.9 kcal/mol-cooperative unit at 20 ºC, which is consistent with my thermal

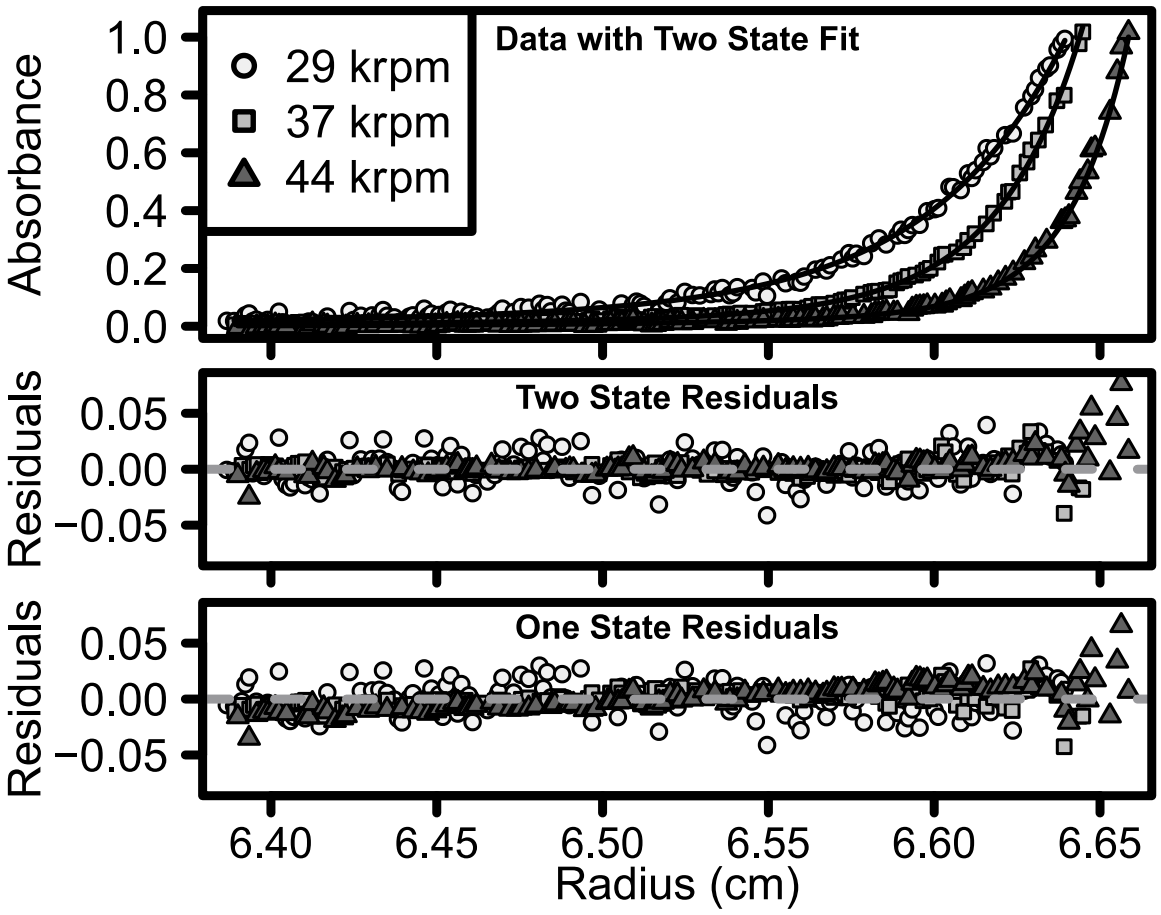


Figure 3.11 – **AUC of the TSG101cc**
The top panel is baselined data from a channel loaded with 9 μM TSG101cc-pyrene. Data are points and two state fits are black lines. The lower two panels are residuals from two and one state fits, respectively. Note that at higher rotor speeds, the residuals are tilted in the one state case. The absorbance shown is for 329, 330, and 329 nm at 29, 37, and 44 krpm, respectively.

measurements (CD at pH 7.2 and 20 ºC ΔG = 23.2 kcal/mol-oligomer, and DSC ΔG = 15.8 kcal/mol-oligomer). The agreement between the DSC, CD and AUC measurements strongly suggests that under native and denaturing conditions, no appreciable amount of oligomers other than tetramer are populated.

3.4.6 Structural Thermodynamic Analysis Using COREX/BEST

To investigate the structural stability of TSG101cc, the crystal structure of the tetrameric complex was analyzed using the COREX/BEST algorithm [149, 420]. COREX/BEST uses the high-resolution structure as a template from which to derive an ensemble of partially unfolded states. The fundamental assumption in the COREX algorithm is that conformational fluctuations can be modeled as local unfolding, and that the folded regions in each state can be represented by the high-resolution structural coordinates. Different states are generated by systematically unfolding all regions (in 8 residue windows) of the protein in all combinations, thus producing an ensemble that ranges from the fully unfolded state, in which all regions are unfolded, to the fully folded state, in which no regions are unfolded. The energetics were determined using the well-established calorimetric parameterization of the enthalpy and heat capacity as described elsewhere [33, 82, 128, 128, 149, 294, 447, 448].

The output of COREX/BEST is a profile of the residue-specific stability of the molecule, which is presented from the N to C terminus Figure 3.12. Because COREX/BEST generates an ensemble, the residue-specific stability is described at every position by the expression:

$$\kappa_{f,j} = \frac{\sum \mathcal{P}_{f,j}}{\sum \mathcal{P}_{nf,j}} \tag{3.18}$$

where the stability at any position j is simply the ratio of the summed probability of all states wherein residue j is in a folded region, to the summed probability of states wherein that residue is in a non-folded region. As such, high stability constants (i.e., $\mathcal{RT} * ln[\kappa_f]$) correspond to regions where the probability of states wherein residue j is folded is high, while lower stability constants are found in regions where the probability of states wherein residue j is not folded is high. Importantly, the stability constants determined by COREX/BEST can be compared with the protection factors obtained from hydrogen exchange experiments under conditions where protection reports on a stability (i.e. under EX2 conditions). The excellent agreement between COREX and a test set of proteins suggests that the regional differences in stability determined from COREX are meaningful [253].

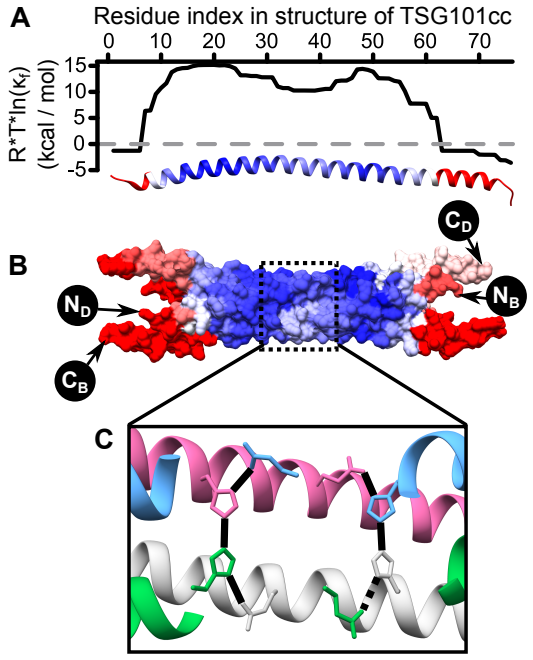


Figure 3.12 – Structural Thermodynamic Analysis of TSG101cc
A) COREX/BEST residue stability analysis for chain A of pdb:3iv1, at pH 7 and 293.15 Kelvin. The *y*-axis is the energy of unfolding a given residue on the *x*-axis. Laid horizontally from N to C-terminus is chain A from the tetramer structure, depicted as a ribbon. A gray dashed line indicates an energy of 0 and was used as the upper limit for the red coloring of the structure. The upper limit for blue coloring was +15. **B)** The molecular surface of the TSG101cc tetramer is colored as in part A. The N and C-termini of chains B and D are labeled to highlight the anti-parallel nature of the tetramer ("N_B" etc.). **C)** Zoom in of the proposed H-bond network is shown as solid lines, except one with relaxed constraints as a dashed line. Chain A = light grey, chain B = green (cut away), chain C = pink, chain D = blue (cut away).

Analysis of the COREX results for TSG101cc reveals two important features.

First, both ends of the homotetrameric coiled coil are predicted to be unstable and largely disordered in solution (parts A&B of Figure 3.12). This is consistent with the fact that the ends of each chain are not in contact with one another in the high-resolution structure, and thus lack the inter-helix stabilization that is characteristic of coiled-coils [368, 404]. Such fraying of coiled-coil ends has been observed in other proteins [84, 156]. Second, the central portion of the coiled-coil is predicted to be highly stable, which is consistent with the high degree of inter-helical surface buried between the individual monomers in the coiled-coil.

A slight decrease in the stability of the central portion of the coiled-coil region is also observed. This decrease is accompanied by an increase in the residue –specific conformational entropy of unfolding of the central portions of the TSG101cc. The increased conformational entropy and decreased stability is in part related to the presence of a glycine (Gly) residue in the center of the helix (Gly268, which is a.a. 40 in part A of Figure 3.12, see Figure 3.13) which increases the conformational entropy of states wherein Gly268 is unfolded, thus destabilizing helical states by 1.0 kcal/mol. Indeed, it is this destabilizing conformational entropic effect that largely accounts for the infrequency of Gly residues within helices of folded, globular proteins [82].

I note the presence of Gly at position 268 is immediately followed by His 269, which is part of an inter-helical H-bond network involving Glu262 of chain A,

His269 of chain B, His269 of chain C, and Glu262 of chain D, all of which are observed in the X-ray structure part C of Figure 3.12 (and which are predicted to be hydrogen bonded using the program Chimera [317]). I also note that the two H-bonding networks per tetramer would, upon dissociation, release one proton per each histidine pair; producing a net release of two protons per tetramer, a value that is in good quantitative agreement with my experimental results. Interestingly, such an H-bond network could also play a role in setting the register and orientation of the coiled-coil, as seen in some other natural and engineered coiled-coils [56, 306].

A more detailed analysis of the COREX/BEST results provides access to the thermodynamic mechanism underlying the homotetramer stability. COREX calculates the unfolding energetics based on changes in solvent accessible apolar (ΔASA_{ap}) and polar (ΔASA_{pol}) surface area between states, from which ΔC_p, $\Delta H_{solvation}$, and $\Delta S_{solvation}$ can be determined [148]. Two other components of the entropy change can also be determined: The conformational entropy, $\Delta S_{conformation}$

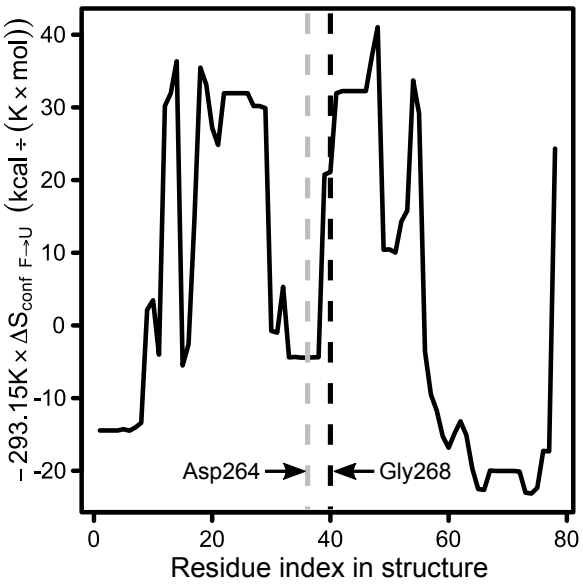


Figure 3.13 – The Energy of ΔS_{conf} on a Per Residue Basis.
The solid black line is the data from COREX. The dashed lines indicate specific residues of interest: Asp264, grey, H-bonds its backbone to Gly268, black.

can be calculated based on backbone and side chain rotomers, as described previously [82, 236], and the translational and rotational entropy, ΔS_{trans}, for each oligomeric state can be calculated given the stoichiometry [295].

Table 3.5 contains the structural thermodynamic analyses of the TSG101cc homotetramer as well as the intermediate oligomeric states, generated in silico by deletion of excess chains. The calculated ΔG_{total} and ΔC_p for the tetramer are in agreement with the values measured experimentally (COREX: 5.2 kcal/mol-monomer and 650-690 cal/K/mol-monomer, respectively), suggesting the algorithm adequately captures the underlying thermodynamic basis of TSG101cc stabilization. Inspection of the detailed mechanism reveals that the major determinant of the stability of the tetramer is the apolar surface buried in the binding interface between helical monomers. Specifically, the burial of apolar versus polar surface area (on a per residue basis) is significantly higher for the tetramer than for the trimer, dimer, or monomer (1.2, 1.6, 2.5 fold more ΔASA_{ap} in the tetramer, respectively). Of note is that the lower order oligomers are predicted by COREX/BEST to be significantly less stable than the homotetramer, which is also consistent with the experimental results.

To estimate the relative stability of the TSG101cc heterotrimer, a structural analysis of the yeast ESCRT-I coiled-coil (pdb:2p22) was also performed [212]. The results reveal a binding interface that is 77% apolar, compared to 75% for the TSG101cc homotetramer (Table 3.6). Analysis of the results also reveals

Parameter	Monomer	Dimer	Trimer	Tetramer
ΔASA_{Apolar} (Å^2)	1331	2105	2776	3382
ΔASA_{Polar} (Å^2)	2645	2834	3077	3345
ΔC_p (cal/K mol)	-88	211	449	652
$\Delta H_{generic}$ (kcal/mol)	75.5	63.0	55.5	50.8
ΔS_{solv} (cal/K mol)	-70	-158	-232	-297
ΔS_{conf} (cal/K mol)	390	420	429	449
$\Delta G_{generic}$ (kcal/mol)	-18	-13.7	-2.26	6.11
ΔG_{trans} (kcal/mol)	0	-0.77	-0.92	-0.95
ΔG_{total} (kcal/mol)	-18	-14	-3.2	5.2

Table 3.5 – Structural Thermodynamic Analysis of the TSG101cc at 20 ºC (293.15 K)
All values are relative to moles of monomer and refer to the unfolding reaction (positive ΔG values indicate a stable folded state). Note that the tetramer ΔG_{total} calculated here is close to the values I measured experimentally (CD = 5.8, DSC = 4.0, AUC = 5.7 kcal/ K*mol-monomer). The translational energy of dissociation was calculated as in [295].

the stability of the ESCRT-I coiled-coil fragment to be similar to the TSG101cc homotetramer, suggesting that under such conditions the relevant equilibrium would involve TSG101cc homotetramer, unfolded TSG101cc monomer and ESCRT-I heterotrimer, with no appreciable accumulation of the other TSG101cc homo-oligomeric species. Consistent with my results, experimental analysis of the full-length ESCRT-I complex revealed that it sediments as a tightly bound, single species in AUC [212, 287].

Finally, to establish context for the TSG101cc analysis presented here, I also compared my results to other oligomeric proteins, including four coiled-coils (thermodynamic data in Table 3.7). I excluded examples that unnecessarily complicated comparisons to TSG101cc. For example, COREX was parameterized

Parameter	Mvb12	Vps23	Vps37	MxV23	MxV37	V^2	Trimer
ΔASA_{Apolar} ($Å^2$)	1244	1824	2604	2329	3368	4154	4462
ΔASA_{Polar} ($Å^2$)	1529	2758	2988	2327	2773	3400	3208
ΔC_p (cal/K mol)	159	99	388	437	785	974	1162
$\Delta H_{generic}$ (kcal/mol)	31.2	67.2	56.3	35.9	27.2	32.7	16.6
ΔS_{solv} (cal/K mol)	-99	-127	-214	-203	-315	-389	-433
ΔS_{conf} (cal/K mol)	290	432	490	377	428	507	467
$\Delta G_{generic}$ (kcal/mol)	-24.9	-22.2	-24.6	-14.9	-6.0	-1.9	6.7
ΔG_{trans} (kcal/mol)	0	0	0	-0.77	-0.77	-0.77	-0.92
ΔG_{total} (kcal/mol)	-24.9	-22.2	-24.6	-15.7	-6.7	-2.6	5.8

Table 3.6 – Structural Thermodynamic Analysis of *S.cere.* ESCRT-I coiled coil at 20ºC
Per unit of monomer. 2p22 also includes the vps23 stability box bound to vps28, and some poorly resolved loops of vps37, which were deleted for this analysis of the coiled-coil alone. The dimer names are: "MxV23"=mvb12—vps23, "MxV37"=mvb12—vps37, "V^2"=vps23—vps37. Vps23 is the TSG101 homologue. The ΔASA_{Apolar} for the binding interface is three times the trimer value, minus each of the monomer values (7714 $Å^2$). The ΔASA_{Polar} for the binding interface is 2349 $Å^2$.

on real proteins, and artificially engineered proteins can have unusual hydrophobic:hydrophilic properties that could skew calculations. Thus, the criteria for inclusion were: naturally occurring, two-state, mesophilic, having published full-thermodynamic analyses ($\Delta H^º$, ΔC_p, and $T^º$), and with structural data available for COREX. As Table 3.8 reveals, the packing of hydrophobic residues is not unique to the TSG101cc. Although my results reveal a similar amount of hydrophobic surface buried in other coiled-coil interfaces, TSG101cc is less stable (i.e., lower T_{ref}) than most other mesophilic, oligomeric proteins, with the difference arising from the accumulation of small differences in polar free energy and conformational entropy. In short, although TSG101cc has clearly

not evolved to populate homotrimeric and homodimeric states, it also does not appear to have evolved to strongly homotetramerize, as both the calculations and my experimental results reveal.

Protein	Total amino acids	Oligomeric state	$\Delta H(T_{ref})$ kcal/ mol	T_{ref} Kelvin	ΔC_p kcal/ K*mol	ΔG (20ºC) kcal/mol	ΔG per residue cal/mol
GroES	679	Heptamer	445.2	372.7	7.2	29	43
HIV protease	198	Dimer	190.5	360.5	3.2	14	71
SecB	620	Tetramer	509.4	358.4	10.6	26	42
p53tet	252	Tetramer	209.2	397.6	1.7	29	115
ROP	126	Dimer	174.7	358.8	2.5	16	127
H2A-H2B	255	Dimer	124.2	337	<2.6>	8.4	33
Arc repressor	106	Dimer	115.7	355.1	1.6	11	104
H3-H4	171	Dimer	86.8	359.6	1.4	7	41
GCN4	112	Dimer	<69.8>	<368.3>	<0.269>	12	107
Melittin	104	Tetramer	147[a]	369[a]	1.8[a]	15[a]	144
TSG101cc	312	Tetramer	226	347.1	2.7	23	74
APC (a.a. 1-55)	110	Dimer	76.5[a]	345.7[a]	0.664[a]	8.8[a]	80

Table 3.7 – **Review of Thermodynamic Data for Oligomeric Proteins in the Literature**
Data are normalized to the moles of oligomer. "< >" indicate values that are averages of the literature data. The ΔH and T_{ref} values are for hypothetical concentrations at which concentration has zero net effect on the stability and $\Delta G = 0$. [a]There was not enough information in the cited texts to readily calculate ΔG from the published fits. I digitized the cited CD data (Engauge Digitizer, Mark Mitchell) and constrained ΔC_p to the value calculated by COREX. I then fit the data, globally where possible, and present my own fitted values alongside the COREX prediction of ΔC_p. The citations for these datasets are in Table 3.8

Protein	Folding ΔASA_{Apolar} (Å^2 and %)	Inteface ΔASA_{Apolar} (Å^2 and %)	Total ΔASA_{Apolar} (Å^2 and %)	Difference ΔASA_{Apolar} % bind−fold	Ref
GroES	23898 (62%)	14028 (59%)	37926 (61%)	-3[a]	[50, 165]
HIV protease	9088 (63%)	1950 (61%)	11038 (63%)	-2[a]	[378, 409]
SecB	30468 (63%)	3675 (66%)	34142 (63%)	3[a]	[87, 309]
p53tet	3000 (59%)	4608 (71%)	7608 (66%)	12[b]	[184, 236]
ROP	4520 (53%)	1877 (73%)	6397 (57%)	20[b]	[35, 386]
H2A-H2B	6547 (56%)	3685 (76%)	10232 (61%)	20[b]	[17, 194]
Arc repressor	2536 (44%)	2675 (68%)	5211 (53%)	24[b]	[49, 55]
H3-H4	6090 (53%)	3740 (78%)	9830 (61%)	25[b]	[17, 195]
GCN4	1270 (40%)	1435 (81%)	2704 (55%)	41[c]	[367, 404]
Melittin	2300 (47%)	3361 (88%)	5661 (65%)	41[c]	[135, 401]
TSG101cc	5324 (33%)	8204 (75%)	13528 (50%)	42[c]	3iv1
APC (a.a. 1-55)	1928 (35%)	1783 (79%)	3709 (48%)	44[c]	[84, 186]

Table 3.8 – **Apolar Surface Area Comparison of Oligomeric Proteins in the Literature**
[a,b,c]Correspond to different groups clustered by k-means analysis of last data column: [a]enzymes/chaperones, [b]mixed structure/function, [c]pure coiled-coils.

Protein	Total amino acids	Oligomeric state	$\Delta H(T_{ref})$ kcal/ mol	T_{ref} Kelvin	ΔC_p kcal/ K*mol	ΔG (20ºC) kcal/mol	ΔG per residue cal/mol
FIS	196	Dimer	2.2	200.6	374.2	23	115.4
CcdB	202	Dimer	2.8	213	371.7	21	103
MetJ	208	Dimer	2.1	156.5	343.2	15	70
N34(L6)C28: SIV	204	Trimer	2.1	165.7	361.3	17	83
LPP56	168	Trimer	0.6	$\sim$168	$\sim$359.8	27	163

Protein	Folding ΔASA_{Apolar} (Å^2 and %)	Inteface ΔASA_{Apolar} (Å^2 and %)	Total ΔASA_{Apolar} (Å^2 and %)	Difference ΔASA_{Apolar} % bind−fold	Ref
FIS	5808 (53%)	2418 (77%)	8226 (59%)	24[b]	[279, 455]
CcdB	10538 (60%)	1403 (77%)	11941 (62%)	17[b]	[31, 256]
MetJ	8971 (55%)	2804 (76%)	11775 (59%)	21[b]	[182, 331]
N34(L6)C28: SIV	5476 (44%)	4615 (81%)	10091 (55%)	37[c]	[178, 259]
LPP56	2668 (28%)	4317 (76%)	6985 (46%)	48[c]	[97, 372]

Table 3.9 – Continuation of Table 3.7 and Table 3.8. Data are presented in the same manner.

3.5 Discussion and Conclusions

3.5.1 Discussion

In this study, I set out to understand the oligomerization of the human TSG101cc in solution. In isolation, I find that the coiled-coil domain significantly populates only two states, a homotetramer and a monomer (the findings are summarized in part A Figure 3.14). My thermodynamic analysis indicates that the intermediate homo-oligomeric states would likely be unstable because hydrophobic packing present in the homotetrameric interactions is missing in the homotrimer/dimer. This has broader implications and immediately suggests that TSG101cc interacts with its binding partners via large, hydrophobic interactions, a supposition that is supported by thermodynamic analysis of the ESCRT-I coiled-coil. Without significant hydrophobic interactions, hetero-oligomers would be unstable relative to the TSG101cc homotetramer.

I also found that the TSG101cc tetramer releases $\sim$2 protons per tetramer upon unfolding. This is consistent with my analysis of the homotetramer structure (pdb: 3iv1). The four histidine residues at position 269 would likely release two protons upon unfolding of the tetramer producing 0.5 protons per monomer. It is unclear why the TSG101cc evolved to be stabilized as pH is lowered; although I do note that cellular pH can vary from $\sim$5.7—7.4 depending on various cellular

states [37, 303, 327, 329, 416, 434]. Determining if these subtle pH changes affect TSG101 function in vivo awaits further study.

Lastly, I find that the tetramer of the TSG101cc is relatively unstable. Compared to the other tetramers (Table 3.7), TSG101cc has the lowest T_{ref} and the second lowest ΔG (20 ºC). This suggests that the unfolded monomer plays a thermodynamically significant role in the functional equilibrium. Even at some of the highest TSG101cc concentrations at pH 7.2, the TSG101cc unfolding curve begins in a physiological temperature range. At more physiological concentrations, the apparent T_m of unfolding is $\approx$37 ºC or lower, suggesting that the reservoir of TSG101cc is a balance of monomer and tetramer. Importantly, because the monomer-tetramer equilibrium is poised near the midpoint, it is most sensitive to concentration changes, meaning that any TSG101 that is sequestered by the ESCRT complex would be readily replenished by a shift in the tetramer to monomer equilibrium. The fact that the stability appears to be tuned by the presence of the conserved residues E262, G268 and H269 (and the corresponding interchain H-bond network), suggests that functional equilibrium is subject to fine-tuning from both pH and concentration changes.

While the TSG101cc by itself is weakly stable, I hypothesize that other interactions in full-length TSG101 may stabilize a homotetramer. It has been known for over a decade that the N-terminal, UEV domain of TSG101 binds a short peptide motif, P(S/T)AP, that is present in many animal and viral proteins that bind

TSG101 [12, 170, 203]. C-terminal to the coiled-coil of TSG101 is a PTAP motif that can bind the TSG101 UEV domain [323, 324]. At the time of its discovery, Pornillos and colleagues proposed that each TSG101 molecule looped around and bound itself—forming an autoinhibited state that other ESCRT proteins would be unable to bind. Here, I can add to this hypothesis by proposing that each TSG101 molecule engages in domain swapping in its tetrameric state (part B of Figure 3.14). Because every other strand of the coiled-coil runs antiparallel to its neighbors, the UEV domain of a neighboring strand could bind the C-terminal PTAP motif of an adjacent strand (part B of Figure 3.14).

Although the model hypothesized in part B of Figure 3.14 is speculative, it does provide context to the results presented here and provides potential avenues for further study. The complex shown in part B of Figure 3.14 would likely be highly stable and autoinhibited. There is at least some *in vivo* evidence for an autoinhibited state of TSG101: mutation of mouse TSG101's PTAP motif caused a precipitous increase in TSG101's association with the ESCRT0 protein, HRS, by yeast two-hybrid [262], and in other yeast two-hybrids TSG101 has been found to self-associate [199, 268, 422]. In its natural context, TSG101 could take advantage of the homotetramer-monomer equilibrium to serve as the TSG101 reservoir, from which the ESCRT-I complex could draw. Local increases in concentration of TSG101 binding partners such as HRS [324] or ubiquitin [393] that both bind to the UEV domain of TSG101, could sequester TSG101 monomers.

In any event, the model proposed in Fig. 7B reveals numerous possibilities for how TSG101 function can be affected by changes in stoichiometry, which can affect both inter and intramolecular interactions, possibly in complex ways. For instance, to bind the TSG101 UEV domain, the HRS PSAP motif may have to displace the TSG101 PTAP motif. However, because it has been shown that the TSG101 PTAP binds its UEV domain approximately 1.5–fold better than does the HRS PSAP motif [324], other domains are likely to play a role in mediating this interaction. Is it the case that binding of ubiquitin, for example, allosterically (or directly) destabilizes binding of the TSG101 PTAP motif, thereby releasing the UEV domain for binding to HRS? Although the answer to this question awaits further study, the fact that the concentration of available TSG101 monomer is within a tunable range suggests that dynamic availability of TSG101 may be a key determinant in HRS function.

3.5.2 Conclusions

In this work, I used solution and structural thermodynamics to define the oligomeric states of the TSG101 coiled-coil *in vitro*. I find evidence that TSG101cc evolved a highly cooperative association event that is stabilized by hydrophobic interactions and acidic pH, which gives me several clues as to how TSG101cc forms its hetero-oligomers not just with ESCRT-I but also with nuclear hormone receptors. The information gained here leads to new hypotheses that could

inform future studies involving full-length TSG101 and its binding partners.

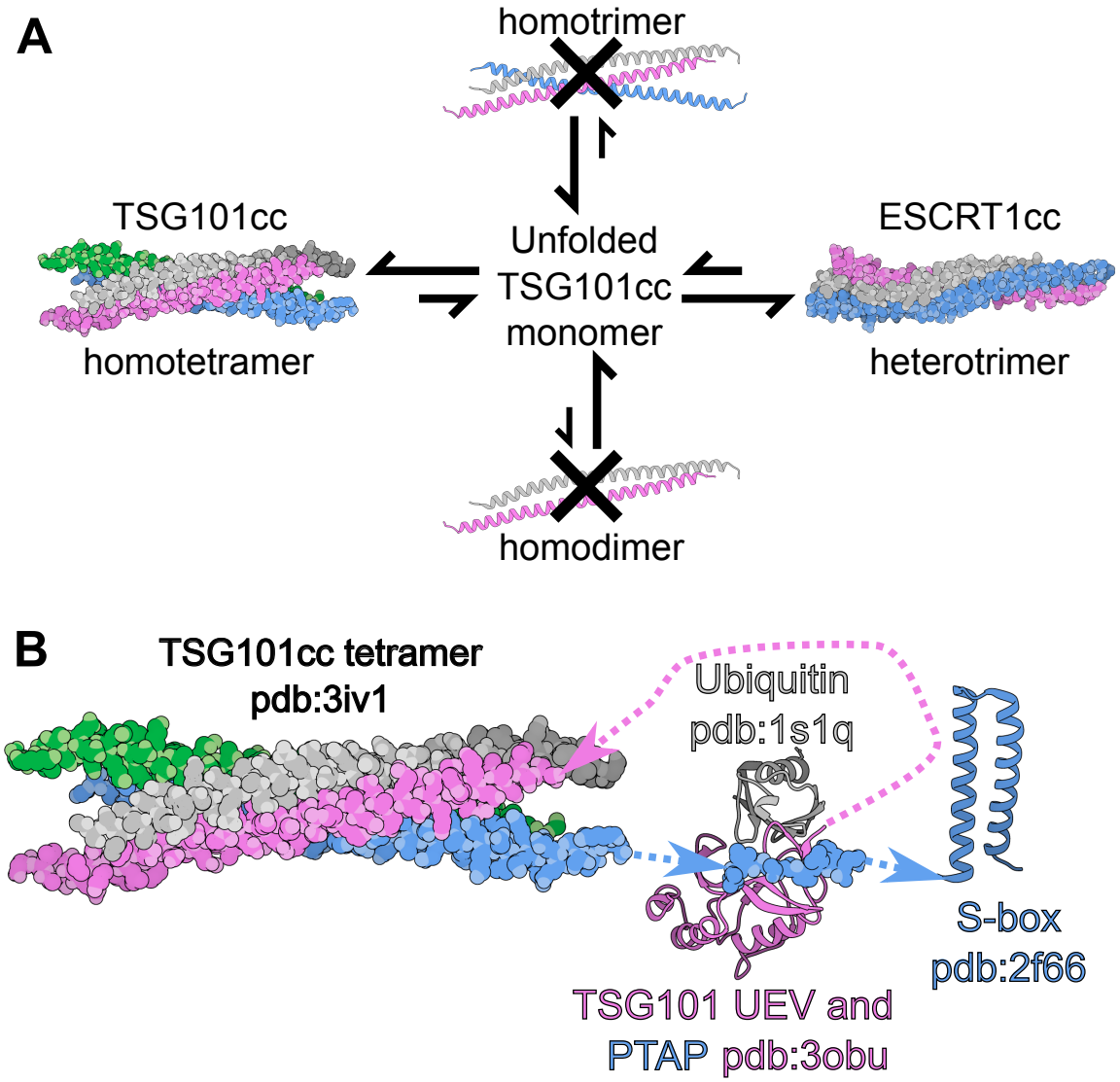


Figure 3.14 – **A) Reaction scheme depicting the main discoveries of this work**. On the left, each chain of the TSG101cc tetramer is depicted as atom spheres with carbons in darker colors than the other atoms (pink, grey, green, blue). At the top and bottom, the TSG101cc is depicted as a ribbon in the same colors. On the right, the TSG101cc homologue, vps23, is blue (vps27 is grey and mvb12 is pink, adapted from pdb:2p22). The arrows are based on structural thermodynamic analyses, presented here, and experimental data presented here for TSG101cc or elsewhere for ESCRT-I [212, 287]. **B) The proposed structural model for the TSG101 tetramer.** Each chain of the coiled-coil tetramer is depicted as in part A of this figure. The N-terminal UEV domain of the pink chain is depicted as a ribbon [170], and was docked [10] to the TSG101 PTAP sequence of the blue chain (the original PTAP in pdb:3obu is the HIV sequence, with little difference). The structure of the UEV domain bound to ubiquitin [393] was aligned with pdb:3obu, and ubiquitin is colored dark grey; note that the PTAP motif does not occlude ubiquitin binding. The C-terminal S-box of the blue TSG101cc chain is depicted as a ribbon [213]. Unstructured regions are shown as dashed lines with arrows going from N to C-termini. There would be three other UEV domain:PTAP interactions not shown here.

3.5.3 Acknowledgments:

VJH is supported by NIH grant GM063747 and NSF grant MCB-0446050. JTW was supported by the NIH training grant 5T32GM007231-39 to the Johns Hopkins University department of biology. Dr. V. Moudrianakis for advice and use of his extremely well kept AUC. Technical advice from: Dr's L. Brand, E. Freire, A. Schoen, M. Rodgers, K. Tripp, C. Sandlin, and numerous other members of the biology and biophysics departments at Johns Hopkins.

3.6 Appendix

3.6.1 Sequences of TSG101cc constructs used

All sequences were verified using Sanger sequencing services from either Genscript or Genewiz. Construct used in CD and DSC, DNA of ORF:

ATGCACCACCACCACCATCATCACCATCACGAAAACTTATACTTCCAGGGTAG
TTCTCTCATCTCTGCGGTCAGTGACAAACTGAGATGGCGGATGAAGGAGGAAATG
GATCGTGCCCAGGCAGAGCTCAATGCCTTGAAACGAACAGAAGAAGACCTGAAAA
AGGGTCACCAGAAACTGGAAGAGATGGTTACCCGTTTAGATCAAGAAGTAGCCGA
GGTTGATAAAAACATAGAACTTTTGAAAAAGAAGGATGAAGAACTCAGTTCTGCTC
TGGAAAAATGA

Translation (TEV cut site marked with *, TSG101cc sequence starts two a.a.

later ”SLI…”): MHHHHHHHHHHENLYFQ*GSSLISAVSDKLRWRMKEEMDRAQAE

LNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK

Construct used in AUC and TCSPC (only change is a cysteine in the N-terminal tag), DNA of ORF:

ATGCACCACCACCACCATCATCACCATCACGAAAACTTATACTTCCAGGGTTG

TTCTCTCATCTCTGCGGTCAGTGACAAACTGAGATGGCGGATGAAGGAGGAAATG

GATCGTGCCCAGGCAGAGCTCAATGCCTTGAAACGAACAGAAGAAGACCTGAAAA

AGGGTCACCAGAAACTGGAAGAGATGGTTACCCGTTTAGATCAAGAAGTAGCCGA

GGTTGATAAAAACATAGAACTTTTGAAAAAGAAGGATGAAGAACTCAGTTCTGCTC

TGGAAAAATGA

Translation: MHHHHHHHHHHENLYFQ*GCSLISAVSDKLRWRMKEEMDRAQA

ELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK

3.6.2 PMIA labeling of cysteine

To verify the extinction coefficient of PMIA, a known mass of PMIA was reacted with a 900 fold excess of L-cysteine in labeling buffer (see methods) with 25% DMSO for one hour. The resulting extinction coefficient was similar to the manufacturer’s measurement in methanol (Setareh standard: 45,000±3,000 cm^{-1} M^{-1} at 343±3 nm; Setareh Lot #14760: 44,700 cm^{-1} M^{-1} at 341 nm; my measure of Lot #14760: 42,000 cm^{-1} M^{-1} at 343 nm).

Chapter 4

The Tumor Susceptibility Gene-101 Coiled-Coil Binds and Allosterically Regulates the Glucocorticoid Receptor

This chapter is in collaboration with two graduate students, James Rives and Marla Tharp, who assisted me with the single molecule force spectroscopy and fluorescence anisotropy, respectively. Professors E. Brad Thompson and Vincent J. Hilser have provided guidance and designed

experiments with me. Professors Schroer and Brand graciously allowed the use of equipment critical to this chapter.

4.1 Abstract

Glucocorticoid Receptor (GR) is a steroid hormone receptor important for the proper development and homeostasis of vertebrates. As with nearly all steroid hormone receptors (SHR), the N-terminus of GR is a large disordered domain that mediates the transcriptional activity of GR. It is known that these disordered domains of SHR's bind transcriptional cofactor proteins, but it is poorly understood how transcriptional cofactors affect the energetic coupling and functional output of SHR's. I addressed these questions using a GR binding partner, Tumor Susceptibility Gene-101 (TSG101), that is involved in endosomal maturation and implicated in transcriptional regulation of several SHR's. Using numerous biophysical and cellular assays, I show that the coiled-coil of TSG101 binds and folds the disordered N-terminus of GR, that this interaction promotes DNA-binding of GR *in vitro*, and that the *in vivo* interaction appears to promote the transcriptional activity of most GR isoforms.

4.2 Introduction

It is becoming increasingly apparent that to understand the molecular mechanism of transcription factors, we must understand the underlying energetics of these proteins [151, 291]. This is an inherently difficult task, as many transcription factors contain large disordered regions that are necessary for transcriptional activity [214, 251]. How disordered domains support the function of transcription factors is still under investigation, but it is known that these domains are enriched for protein binding sites and are frequently expressed as alternative isoforms [61]. Alternative isoforms can disrupt the binding of some transcriptional cofactors [61], but less studied is how alternative isoforms could change the underlying coupling energies between domains of a protein [291].

As others have shown for the adenoviral protein E1A and the Hilser lab has shown for glucocorticoid receptor, alternative isoforms can drastically change the allosteric coupling between domains of a protein, even when some of the domains are disordered [106, 242, 243]. This goes against the classical view of allostery as being a bond path or rigid body motion through a protein [91, 226, 315]. Instead, allostery depends upon the ensembles of folded and unfolded states and how folding of one domain affects the stability of the other domains (allosteric coupling) [150, 289].

To understand how allosteric coupling affects transcription, I have turned

to the steroid hormone receptor family of proteins, specifically glucocorticoid receptor (GR). GR regulates a wide variety of processes in the human body including inflammation, cell death, stress and immune responses [77, 267, 354, 407, 421, 456]. GR is typical of the steroid hormone receptor family: it has a disordered N-terminal domain (NTD), a DNA-binding domain (DBD), and a ligand binding domain (LBD). Figure 4.1 lays out these domains and describes some of their functions, binding partners, and translational isoforms (A, B, C1, C2, C3, D1, D2, D3).

The eight translational isoforms of GR progressively chop off the NTD and have drastic effects on the transcriptional activity of the protein [260]. Regardless the presence of the LBD[1], cutting the NTD down to the C3 isoform increases the transcriptional activity of GR, while removal of the C3-NTD destroys GR's transcriptional activity [243, 260]. The Hilser lab has previously measured the folding energy of GR NTD isoforms and found that the C3-NTD more readily folds than the A-NTD [242]. Furthermore, the energy of folding a given GR NTD correlated well with transcriptional activity, and both could be deconvoluted in terms of folding free energies (ΔG) and coupling energies (Δg) between the

[1]The LBD of GR is dispensable to the core functions of GR, and here I use a GR construct lacking the LBD to simplify my thermodynamic analysis. In a similar manner, there are two GR splice isoforms that lack functional LBD's, GR-β and P. Under normal circumstances, GR-β and P are expressed at levels about one-hundredth or one-tenth that of the main GR-α splice isoform, respectively [350]. However, the truncated GR's are upregulated in some disease states: β in both Cushings syndrome and drug-resistant asthma, and GR-P in some cancers (e.g. some forms of acute lymphoblastic leukemia and non-Hodgkin lymphoma) [86, 134, 240]. Understanding the thermodynamic coupling within GR's first two domains will ultimately help us understand the receptor as a whole, but it could immediately impact our understanding of the disease-associated GR-β and P.

domains [243, 289].

The Hilser lab's quantitative model predicts how different perturbations (ex. mutations or binding events) will affect GR function. Because the C3-NTD is more stable than the longer isoforms, I predicted that a transcriptional cofactor would bind the C3 isoform better than the A isoform. However, as described in this chapter, the actual situation appears to be more complex. There are perhaps hundreds of transcriptional cofactors that interact with GR [255], and my choice of transcriptional cofactor was based on several criteria: **1)** known to bind the GR NTD, **2)** ease of purification, and **3)** no requirement for post-translational modifications in the binding site. This led me to tumor susceptibility gene-101.

Nearly two decades ago, it was shown that Tumor Susceptibility Gene-101 (TSG101) could alter transcriptional activity of nuclear hormone receptors [432], and in the intervening time there has been a smattering of reports supporting the initial observations [153, 293, 392]—while most TSG101 research has focused on its conserved role in endosomal maturation (the Endosomal Sorting Complex Required for Transport, ESCRT, see Chapter 2 and: [142, 241]).

To date, the role of TSG101 in transcription is not understand .[2] Indeed, it is not clear if TSG101 forms a transcriptional complex or if it simply regulates receptor degradation or post-translational modifications [63, 299]. Conflicting

[2]It is worth noting that the yeast version of TSG101 has no known nuclear roles (Prof. Wendland, communication). To define the evolutionary course of TSG101, I have aligned the human TSG101 sequence with ten of its homologues in Figure 4.22 in Appendix. A possible nuclear localization sequence appears to be relatively conserved in vertebrates; thus, TSG101 was likely neo-functionalized millions of years ago.

reports of TSG101 activating or repressing transcription remain unexplained [63, 392]; though, some of the differences may be a result of other transcription factors being recruited. In one example, the combination of TSG101 and Daxx repressed transcriptional activity of GR more so than either protein alone [293]. Table 4.1 shows a comparison of literature results with regards to TSG101's nuclear localization, transcriptional effects, and potential binding partners. Also see Figure 4.1 for an explanation of the TSG101 domains and functions. Of particular interest is the coiled-coil domain of TSG101, which was shown to bind GR's NTD in a yeast two-hybrid assay [153].

In the work presented here, I found that the TSG101 coiled-coil (TSG101cc) can bind and fold the GR NTD, regardless of the GR isoform. Overexpression of TSG101 in U2-osteosarcoma cells increased GR's activity, as long as the C3 portion of the NTD was intact, and this *in vivo* effect correlated with *in vitro* experiments where TSG101cc allosterically promoted GR's DNA binding affinity via mechanisms that appear to differ between different GR isoforms. I propose that TSG101 can form an active transcriptional complex with GR.

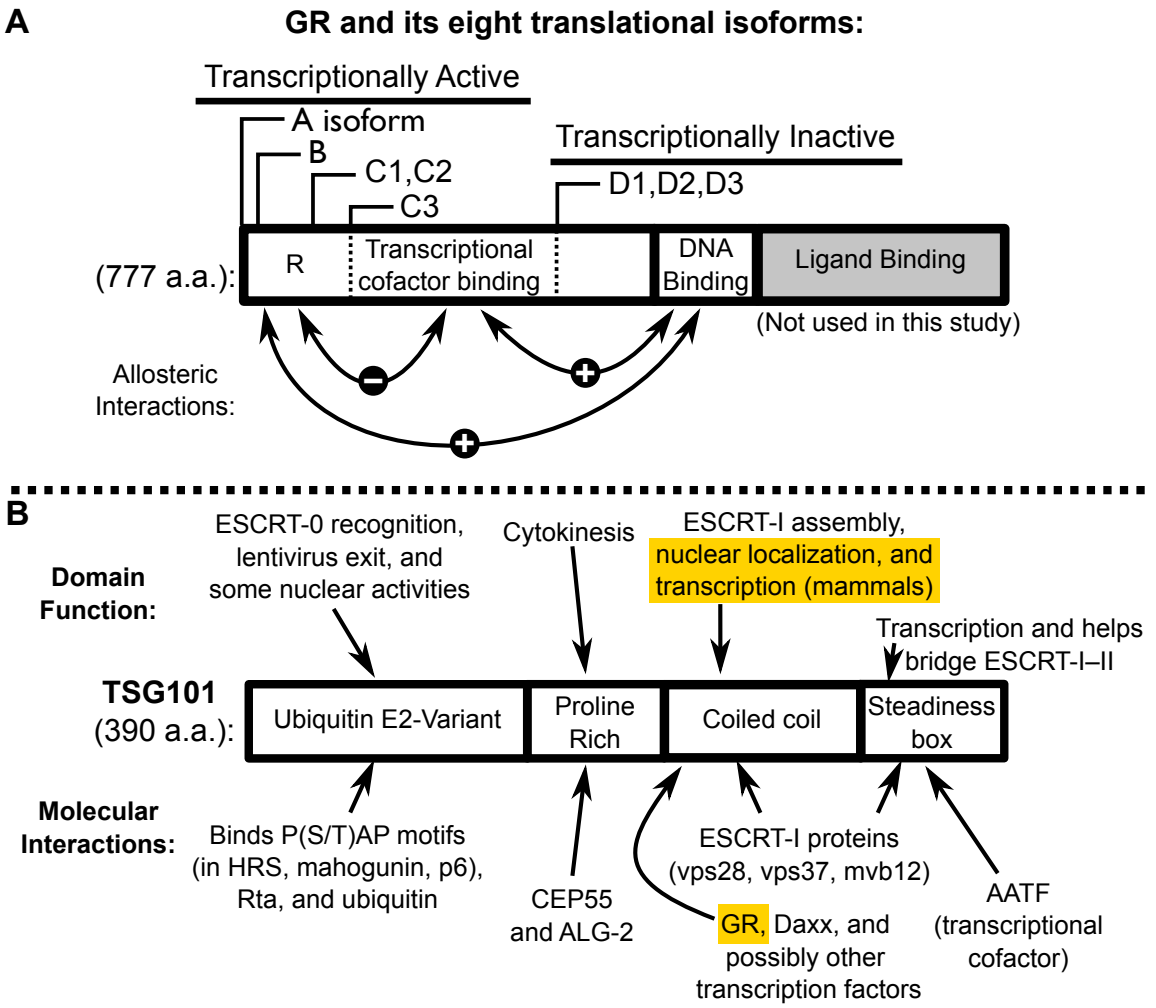


Figure 4.1 – The Domains and Functions of GR and TSG101
Domains are given as boxes oriented from N- to C-terminus, scaled with respect to the size of each domain. **A)** The eight translational isoforms of GR are indicated. Some have nearly adjacent translational start sites and are indicated with just one line (e.g. C1, C2). The longest NTD isoforms contain a regulatory (R) region that stabilizes the DBD and destabilizes the cofactor binding region (depicted as arrows with +/– signs). The shown allosteric interactions produce a bistable system [243]. The NTD region that binds transcriptional cofactors is also called activation function 1 (AF1) in the literature. AF1 has been shown to bind: TSG101, Med14, TBP, CBP, Ada2, and many other cofactors [11, 153, 216, 425]. **B)** TSG101 domain functions are listed at the top, and the coiled-coil's nuclear involvement is highlighted. Examples of protein binding partners are listed at the bottom. **References:** [63, 74, 117, 153, 199, 203, 212, 234, 262, 293, 393, 432]

Cell Line	TSG101 in Nucleus?	Protein Interaction(s)	Transcriptional Effect	Ref.
Rat-1 CV-1	Y	AATF	+ AR	[63]
CV-1	N/A	p300	– * AR, ER, TR	[392]
293T	Y	GR, Daxx	– GR	[293]
NPC-TW01	Y†	Rta (viral)	+ Rta	[74]
Hela	N/A	GR	– GR	[153]
T24 Saos-2	Y‡	N/A	N/A	[459]
3T3 Hela	Y‡	N/A	N/A	[449]
COS-1	N/A	N/A	– AR, GR, ER, TR, RAR, VDR	[432]
U2OS	Y	GR	+	Here

Table 4.1 – Compilation of Literature Data on TSG101's Transcriptional Effects and Interactions
Localization of TSG101 in these studies was determined by either immunofluorescence or fluorescent protein tagging. Protein-protein interactions were determined by W-blotting or yeast two-hybrid. Transcriptional effects were determined by enzyme expression (often luciferase or CAT) using artificial or viral promoters. **Key:** Y = Yes, N/A = Not tested/applicable, + = Activator, – = Repressor. **Cell lines:** Rat-1 = rat fibroblasts, CV-1 and COS-1 = African green monkey kidney, 293T = human embryonic kidney, NPC-TW01 = human nasopharyngeal carcinoma, Hela = human cervical cancer, T24 = human bladder carcinoma, Saos-2 = human osteosarcoma, 3T3 = mouse embryonic fibroblast. **Addenda:** *Activation was seen with a Gal4 fusion to the TSG101 UEV domain. †TSG101 was only visibly in the nuclei of cells that were expressing the Epstein-Barr virus protein, Rta. ‡These studies observed the cell cycle-dependent localization of TSG101 (nuclear to varying degrees in several stages).

4.3 Methods

4.3.1 Cloning of Plasmids Expressing Human GR or TSG101

All *E. coli* expression was achieved with the pJ411 plasmid of DNA2.0, and the cloning of most constructs was described previously [242, 243, 438]. For the largest GR NTD constructs (A = a.a. 1—420, C3 = a.a. 98—420), the coding sequence was tagged with a 9x His tag on both the N and C-termini along with TEV cut sites to remove the tags. The NTD-DBD constructs of GR (A = a.a. 1—525, C3 = a.a. 98—525) were tagged on just the N-terminus with a 6x His-Thrombin sequence (expression problems were observed with TEV or longer His tags). The open reading frame of the TSG101 coiled coil (a.a. 229—304 or —339) was tagged with a 9x His tag followed by a TEV site and a serine on the N-terminus. The TSG101 construct from a.a. 229 to 339 was produced by using PCR to sequentially add the extra sequence to the 229—304 construct that was described previously [438]. Note that the 339 construct extends about half-way into the steadiness box of TSG101 and was the minimal sequence for binding GR in a yeast two-hybrid assay [153].

I also worked with smaller fragments of the GR-NTD to discern the exact sequence that interacts with TSG101. These smaller fragments consist of a region historically called the activation function-1 or the τ1 region (AF1, a.a.

77—262) that is known to bind transcriptional cofactors [11, 75, 153]. One of my constructs was the bait used in a yeast two-hybrid study by Hittelman et al., called AF1$^{107\text{-}237}$ hereafter [153] (Hittelman AF1 = a.a. 107—237) and the other is the AF1 core (a.a. 187—244) that binds TATA-binding protein [11]. For bulk experiments, the AF1 core construct was tagged with a single 9x His-TEV site on its N-terminus (DNA2.0). For single molecule force spectroscopy, the AF1 core and the Hittelman AF1 were tagged with a 9xHis-TEV-Avi tag on the N-terminus (Avi = LNDIFEAQKIEWHE, receives a biotin on the K) and a ybbR tag on the C-terminus [189, 252, 290, 452] (Ybbr = DSLEFIASKLA, receives CoA on the first S). The C3-NTD single molecule construct was similar, except with a 6xHis tag. All the single molecule force spectroscopy constructs were cloned from existing pJ411 expression plasmids using Gibson cloning.

Besides luciferase, all U2OS expression was achieved with the pJ603 plasmid of DNA2.0, and the cloning of the GR constructs was described previously ([243]; A = a.a. 1—525, B = a.a. 27—525, C1 = a.a. 86—525, C2 = a.a. 90—525, C3 = a.a. 98—525, D1 = a.a. 316—525, D2 = a.a. 331—525, D3 = a.a. 336—525). Note that the pJ603 plasmid uses a constitutive promoter (cytomegalovirus) to express the inserted gene. DNA2.0 generated the full-length TSG101 expression plasmid with codon optimization for human cell expression. Sub-cloning of the TSG101 pJ603 plasmid was achieved with PCR based amplification of either

the coiled-coil (a.a. 228–304)[3] or excision of the same region. To make the pJ603 TSG101cc plasmid, I used primers designed for Gibson cloning of the existing full-length TSG101 plasmid and insertion of an N-terminal methionine. Gibson cloning proceeded by manufacturer specifications (NEB). Excision of the coiled-coil was done with PCR primers that amplified out the coiled-coil and added BamHI and KpnI sites to the N- and C-terminal ends of the cut. A linker oligo (expressing GSGGGSGGSGGSGGSG) with BamHI and KpnI sites was then inserted using ligation.

To tag TSG101 with Tandem Tomato fluorescent protein (tdTom), I used PCR to amplify full-length TSG101 from pJ603 and add a BamHI site at the 3' end. This was then linked to a double stranded oligo that codes for a 13 a.a. linker (GSGSGSGGSGSGT) and has BamHI on its 5' end and KpnI on its 3' end. TdTom was amplified from its plasmid (pCMV-tdTom from Clontech) using primers that added KpnI and XhoI to the 5' and 3' ends, respectively. All of the above fragments were then assembled by ligation into the pJ603 plasmid and sequence verified.

See appendix for sequences.

4.3.2 Protein Expression and Purification

Expression and purification of TSG101cc and the various GR constructs used here has largely been described before [242, 243, 438]. All recombinant protein

[3]Note that my TSG101cc pJ603 construct starts at a.a. 228 as opposed to 229 of my *E. coli* construct. This was to form a more canonical Kozak consensus sequence.

expression was done using the pJ411 vector from DNA 2.0 (see above). Each plasmid was transformed to BL21 (DE3) pLysS cells and plated to kanamycin plates (30 µg/mL). Individual colonies or glycerol stocks were grown overnight at 37º C in 3 mL of LB per liter of desired expression volume. The next morning, cultures were transferred to flasks with 10 mL of LB per desired liter of expression volume, grown for 3 hours at 37 ºC, and transferred to the final expression volume. Growth proceeded around room temperature for about 4 hours before reaching an $O.D._{600}$ ≈0.4, then temperature was reduced to 16º C, and induction occured at $O.D._{600}$ ≈0.6 with 1 mM IPTG. After overnight expression, cells were pelleted by centrifugation (3580 rcf, 10 mins, 4º C), washed with PBS, and repelleted.

The GR NTD alone (either A or C3 isoform) and TSG101cc were prepped from inclusion bodies, often by directly lysing the cells with ∼15 mL Gdn buffer per liter of expression (6 M Gdn HCl, 500 mM NaCl, 20 mM imidazole, 20 mM Tris base, at pH 8). The two domain constructs of GR were prepped under native conditions (100 mM Na_2HPO_4, 400 mM NaCl, 20 mM imidazole, 1 mM TCEP added fresh, ∼5 µL DNAse I, at pH 8). Lysates were cleared by centrifugation (1 hour, 26200 rcf, 4º C) and filtered (borosilicate glass, 1 µm), before running over Ni-NTA equilibrated with the appropriate lysis buffer. Columns were washed with lysis buffer until a stable absorbance was detected and then eluted with 400 mM imidazole.

Nickel purified protein was then subjected to tag cleavage by Tobacco Etch

Virus protease (TEV) or thrombin. Cleavage buffer conditions were either 20 mM HEPES, 40 mM NaCl, and 1 mM DTT at pH 8 (TEV) or 20 mM K_2HPO_4, 250 mM NaCl, 1 mM TCEP, and pH 8 (thrombin). TSG101cc and the GR NTD constructs were next purified by anion exchange, and after overnight cleavage their respective dialysis buffers were changed to 2% anion buffer B (20 mM Tris, ∼20 mM NaCl, 5% glycerol, pH 8). Protein was then 0.2 µm filtered and run on a 5 mL HiTrap Q column (General Electric) using a GE Äkta Pure FPLC system. Neither anion nor cation exchange were effective for the NTD-DBD constructs, so they were simply run over benzamidine-agarose and then had PMSF (0.2 M in DMSO) added to 0.2 mM to remove thrombin and inhibit protease activities, respectively. All constructs were then run back over Ni-NTA that was equilibrated with 0 M imidazole, native buffer. Because the GR constructs often had degradation products, I found that SEC was necessary for reproducible results and the highest purity (Hiload 16/600 column packed with Superdex pg-75 from GE; 20 mM HEPES, 200 mM NaCl, 1 mM TCEP at pH 7.2). Chromatography fractions were run on SDS-PAGE and stained using the coomassie “silver” protocol Candiano2004. For experimentation, proteins were dialysed overnight to their respective buffers (below). I must emphasize that purification of the NTD-DBD construct of GR is not trivial and must be done as quickly as possible to prevent oxidation and artifactual results, even in the presence of TCEP.

For fluorescence anisotropy of TSG101cc, a cysteine was engineered into the expression tag (replacing a serine after the TEV site) and protein was labeled with pyrene-iodoacetamide as described previously [438].

See Figure 4.2 for examples of GR and TSG101 protein purity.

4.3.3 Protease Protection, Cross-Linking, and Mass Spectrometry

Protease digestions were carried out in both the fluorescence anisotropy buffer and the circular dichroism buffer (see below), with no observed differences.[4] Saturation of GR required at least a thirty-fold excess of TSG101cc. To control for the total amount of protein present in each digest, the ratio of trypsin and targeted protein was held constant at 1000 µg protein : 1 µg trypsin, except where noted. All digests occured at room temperature after allowing the protein solutions at least 20 minutes to equilibrate. Each timepoint was taken by quenching an aliqout of reaction in Laemli sample buffer and heating at 95º C for ∼5 minutes.

Three cross-linkers were tested as follows, with their linker lengths in parentheses: BS3 (11.4 Å), DST (6.4 Å), and EDC + sulfo-NHS (0 Å) from ThermoScientific or ProteoChem. BS3 and EDC proved to be wholey unusable in these circumstances, either because of too much or no cross-linking, respectively. The

[4] My anecdotal observations suggest that as long as a buffer has a somewhat low salt concentration (< 200 mM) and a pH ∼7, binding of GR and TSG101cc can be observed.

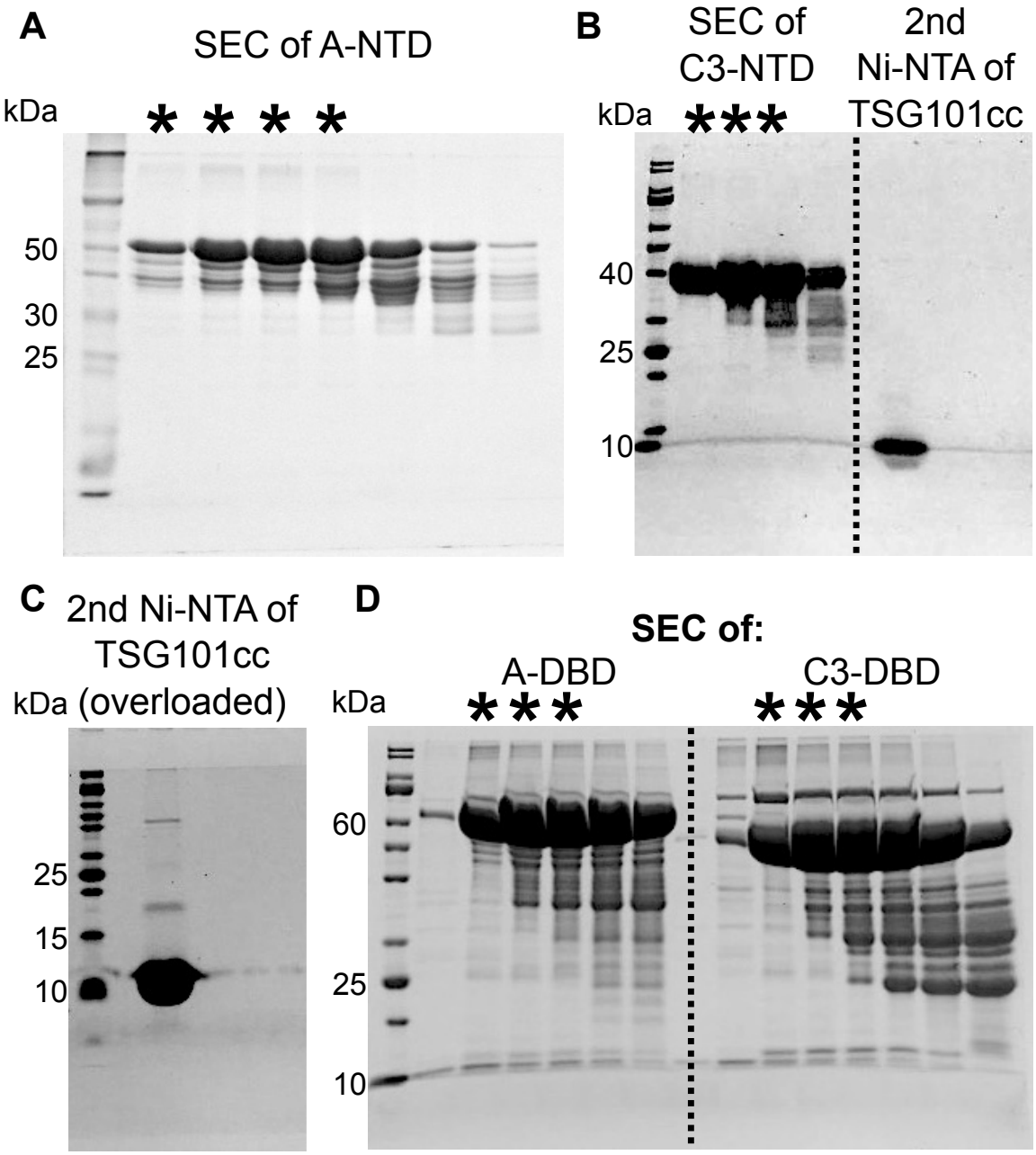


Figure 4.2 – **GR and TSG101 Protein Purity**
All of these gels were stained with coomassie and the images have been grayscaled to give better contrast. NEB ladder marks are on the left of each image. Asterisks indicate fractions that were taken for experimental analysis (ideally 90% purity or better). **A)** SEC of the A-NTD of GR on a 12% PA gel. **B)** SEC of the C3-NTD of GR (left) and flow-thru from the 2nd pass of TSG101cc over Ni-NTA (after tag removal). This is a 12% PA gel and TSG101cc runs with the loading dye. **C)** A 15% PA gel overloaded with TSG101cc to show minor contaminants and the fact that TSG101cc runs slightly high on gels (should be 9040 Da and actual mass is correct by MALDI). **D)** SEC fractions for the A-DBD (left) and C3-DBD (right) of GR. More of the C3 fractions were run on the gel and the reader will note highly degraded GR fragments left the SEC column after the main GR peak.

rest of this manuscript only refers to DST cross-linking. The DST cross-linking buffer was the same as the fluorescence anisotropy buffer (below). Initial reactions were carried out according to the manufacturer protocols, but because TSG101cc has a large number of lysines (11 out of 78 a.a.) I observed an excessive amount of TSG101cc self cross-linking. After lowering the concentration of DST to about one-sixth or one-twelfth the recommended amount (0.5 mM used here) I was able to detect a GR—TSG101cc cross-link without TSG101cc background.

All cross-linking lasted about 5 minutes at room temperature ($\sim$20º C). After quenching each cross-linking reaction with > 10 mM Tris for 15 minutes (1 M stock at pH 7.5), each reaction was reduced with an excess of > 10 mM TCEP for an hour (1 M stock with HEPES at neutral pH), and then cysteines were alkylated with at least 10 mM iodoacetamide for an hour at room temperature (freshly solubilized in experimental buffer and kept dark).

SDS-PAGE was used to separate products of protease protection and protein cross-linking. After staining gels with coomassie [68], desired bands were cut out and diced into $\sim$1 mm slices. A blank part of each gel was also cut out as a negative control. Destaining was achieved with 50 mM NH_4 HCO_3 in 50% acetonitrile for at least 30 minutes. Stain was discarded and the slices were then desiccated with 100% acetonitrile for about 30 minutes. After removal of acetonitrile solution, cold solutions of trypsin (6.25 ng/µL final; 0.1 µg/µL stock

in 50 mM acetic acid) or trypsin + chymotrypsin (6.25 ng/µL final; 1 µg/µL stock in 1 mM HCl) were diluted into 50 mM NH_4 HCO_3 and incubated with the gel slices for an hour at 4º C. Before overnight incubation at 37º C, volumes were adjusted with extra protease or 50 mM NH_4 HCO_3 to ensure complete liquid coverage of the gel slices.

Digestions were mixed with α-cyano-4-hydroxycinnamic acid (CHCA) in 50% acetonitrile with 0.5% tri-fluoro-acetic acid at ratios between 1 part digest : 1 to 4 parts CHCA. Protease protection digests were directly plated in this manner for matrix assisted laser desorption ionization mass spectrometry (MALDI). To increase signal from barely visible cross-link products, I extracted and vacu-fuged the cross-linking digests. For extraction of gel slices, the supernatant of each overnight digest was removed to a fresh lo-bind tube (Eppendorf), and the gel slices were extracted for 15 minutes at 37º C with 67% acetonitrile and 1.7% formic acid (2 parts ACN : 1 part 5% FA). The tubes were then sonicated in an ice bath, and the supernatants were combined and vacu-fuged until ∼20 µL were left. Tubes were covered with perforated parafilm during vacu-fuging to minimize any contamination.

MALDI mass spectrometry proceeded largely as described before [438]. A Bruker AutoFlex III mass spectrometer was used in reflectron mode. The m/z values were standardized using the ProteoMass MALDI calibration kit (Sigma), and further calibrated using TSG101 or GR peaks that were highly consis-

tent. Cross-link ion masses were predicted given the following criteria: trypsin and/or chymotrypsin were allowed to potentially miss up to five cut-sites and carbamidomethylation (iodoacetamide) was left as a variable modification. Data were also analyzed for up to three mono-links and any loop links, with only TSG101cc reproducibly yielding a monolink (see Table 4.9 in Appendix). Data were hand curated and compared to the following negative controls: a blank piece of gel (also received proteases), DST reacted TSG101cc and its oligomeric cross-links, DST reacted C3-NTD, unreacted C3-NTD, and unreacted TSG101cc.

4.3.4 Fluorescence Anisotropy

Preliminary experiments used a variety of buffers, and to enable experiments involving the addition of DNA, the following buffer was used for all fluorescence anisotropy experiments shown here: 13.2 mM HEPES, 16 mM K Acetate, 67.2 mM NaCl, 4.2 mM $MgCl_2$, and 0.84 mM TCEP all pH'd to 7.56 using 10 M KOH. It is difficult to purify large amounts of any GR construct that includes the disordered NTD. In order to maximize the high concentration end of my experiments, all the titrations presented here are dilution experiments: The fluorophore (pyrene-TSG101cc or 6FAM-DNA) was diluted into a concentrated solution of GR and the GR was then diluted out with a solution of fluorophore, with the concentration of fluorophore being kept the same throughout. The dilution solution was actively held at 20º C during the experiments to prevent

thermal disequilibrium.

All experiments were at the very least duplicated with independent protein preparations, and all replicate data sets were fit globally to produce the presented values. Equilibration times were varied to check for thermodynamic equilibrium, and because no obvious effects were discerned from 1 to ∼30 minutes, most of the data presented were collected with 2 minute incubations between measurements. The only sensitive incubation step was the first one, which needed to be about 20 minutes long to establish thermal equilibrium.

Fluorescence anisotropy of pyrene-labeled TSG101cc (always at 76 nM) was obtained with a SLM-Aminco SLM-48000S spectrofluorometer at 20º C. Samples were placed in a 1 cm, square quartz cuvette with a stir bar and excited at 345 nm with a 4 nm bandwidth. Emission was detected at 377 nm with a 8 nm bandwidth and 25 averages per measurement.

GR usually forms dimers on DNA by binding adjacent sites [263]. Such DNA elements are called GR response elements (GRE). For anisotropy of 6-FAM DNA, I used both the canonical GRE and a half-site DNA (single GR binding site). The canonical GRE used was: 5’ 6-FAM-TGCC**AGAACA**GAG**TGTTCT**GACG 3’ and its unlabeled reverse complement [163] (ordered from IDT). The individual GR binding sites are highlighted in bold font. The GR half-site DNA was: 5’ 6-FAM-CGGC**AGAACA**GGACGCG 3’ and its reverse complement. Both constructs were annealed at 50 µM using the IDT annealing buffer (100 mM K Acetate, 30

mM HEPES at pH 7.5) and heating at 95º C for 1 minute followed by bench-top cooling for 1 hour.

Fluorescence anisotropy of 6-FAM labeled DNA (always at 10 nM) was obtained with an Aviv spectrofluorometer at 20º C. Samples were placed in a Starna sub-micro cuvette (cat. no. 16.100F-Q-10/Z15). Excitation was at 496 nm with a bandwidth of 5 nm, and emission was observed at four wavelengths (524, 520, 516, and 514 nm) with a bandwidth of 4 nm and 5 seconds of averaging. Anisotropies from the four emission wavelengths were averaged for final data analysis. Data were fit using equations of the following form:

$$< \alpha >= (a + b * ln[\mathcal{GR}]) * \mathcal{P}_{\mathcal{B}ound}(\mathcal{GR}) + (c + d * ln[\mathcal{GR}]) * \mathcal{P}_{\mathcal{U}nbound}(\mathcal{GR}) \quad (4.1)$$

where $< \alpha >$ is the experimental observable, "a, b, c, d" are all parameters dictating the anisotropy baselines, and $\mathcal{P}$ is the probability of some listed state as a function of GR concentration (note that $\mathcal{P}_{\mathcal{U}nbound} = (1 - \mathcal{P}_{\mathcal{B}ound})$). In all fitting, the goal was to minimize systematic residuals, without overfitting the data. For this reason, if logarithmic baselines did not improve the fit greatly or caused obvious overfitting, the b and/or d parameters were set to zero.

Derivation of the $\mathcal{P}_{\mathcal{B}ound}$ proceeded from first principles. Several potential derivations (single-site, two-site independent or dependent, etc.) were used to test each data set's compliance with the proposed binding mechanism. Only

the equations used are discussed here. In the ideal case, the fluorophore being observed is at a concentration much less than the K_d value. This was the case for binding of pyrene-TSG101cc to the GR NTD. Thus:

$$\mathcal{P}_{\mathcal{B}ound} = \frac{\mathcal{GR}/\mathcal{K}_{\mathcal{d}}}{1 + \mathcal{GR}/\mathcal{K}_{\mathcal{d}}} \tag{4.2}$$

where GR is assumed to be the total concentration of GR in molar and the denominator is the partition function.

For binding of GR NTD-DBD constructs to the full-length GRE, the GRE was often at concentrations close to the apparent K_d. For this reason, all fitting of 6FAM-GRE data, both full and half-site, was done with the full polynomials describing binding as a function of both GR and GRE. For the half-site GRE, the single site binding equation is:

$$\mathcal{P}_{\mathcal{B}ound} = \frac{\mathcal{GR} + \mathcal{GRE} + \mathcal{K}_{\mathcal{d}} - \sqrt{(\mathcal{GR} + \mathcal{GRE} + \mathcal{K}_{\mathcal{d}})^2 - 4 * \mathcal{GR} * \mathcal{GRE}}}{2 * \mathcal{GRE}} \tag{4.3}$$

where GRE is the total concentration of GRE in molar.

Binding of GR to the full GRE is a two-site binding problem. It was found that any equation including a singly bound GRE would produce wild aberrations in the fits; thus, singly bound GRE does not exist at equilibrium and was excluded from the fitting equation. This is the definition of an all-or-none binding process

with a very high and positive cooperativity. The equation with its derivation first:

$$\mathcal{K}_{d} = \frac{\mathcal{GRE}_{unbound} * \mathcal{GR}^{2}_{unbound}}{\mathcal{GRE}x\mathcal{GR}_{2}}$$

$$\mathcal{GRE}_{total} = \mathcal{GRE}_{unbound} + \mathcal{GRE}x\mathcal{GR}_{2}$$

$$\mathcal{GR}_{total} = \mathcal{GR}_{unbound} + 2 * (\mathcal{GRE}x\mathcal{GR}_{2})$$

Substitute into $\mathcal{K}_{d}$ and solve for the bound complex concentration ($\mathcal{GRE}x\mathcal{GR}_{2}$).

$$0 = (\mathcal{GRE}_{total} - \mathcal{GRE}x\mathcal{GR}_{2}) * (\mathcal{GR}_{total} - 2 * \mathcal{GRE}x\mathcal{GR}_{2})^{2} - \mathcal{GRE}x\mathcal{GR}_{2} * \mathcal{K}_{d} \quad (4.4)$$

Solve for the positive, real root of Equation (4.4) to get $\mathcal{GRE}x\mathcal{GR}_{2}$, then divide by the total concentration of GRE to get the probability of GRE bound, shown here with $\mathcal{GRE}_{total}$ and $\mathcal{GR}_{total}$ given simply as $\mathcal{GRE}$ and $\mathcal{GR}$:

$$\mathcal{P}_{Bound} = \frac{1}{\mathcal{GRE}} * \left(\frac{\mathcal{GR} + \mathcal{GRE}}{3} + \left[16 * \mathcal{GR} * \mathcal{GRE} - 4\mathcal{GR}^{2} - 16\mathcal{GRE}^{2} + 12\mathcal{K}_{d}^{2}\right] \div \left[24 * \mathcal{A}\right] - \frac{1}{6} * \mathcal{A} \right)$$

Where $\mathcal{A}$ is:

$$\left\{ \mathcal{GR}^{3} - 6 * \mathcal{GR}^{2} * \mathcal{GRE} + 12 * \mathcal{GR} * \mathcal{GRE}^{2} - 8 * \mathcal{GRE}^{3} + 9 * \mathcal{GR} * \mathcal{K}_{d}^{2} + 9 * \mathcal{GRE} * \mathcal{K}_{d}^{2} + 3\sqrt{3} * (\mathcal{GR}^{4}\mathcal{K}_{d}^{2} - 6 * \mathcal{GR}^{3} * \mathcal{GRE} * \mathcal{K}_{d}^{2} + 12 * \mathcal{GR}^{2} * \mathcal{GRE}^{2} * \mathcal{K}_{d}^{2} - 8 * \mathcal{GR} * \mathcal{GRE}^{3} * \mathcal{K}_{d}^{2} + 2 * \mathcal{GR}^{2} * \mathcal{K}_{d}^{4} + 10 * \mathcal{GR} * \mathcal{GRE} * \mathcal{K}_{d}^{4} - \mathcal{GRE}^{2} * \mathcal{K}_{d}^{4} + \mathcal{K}_{d}^{6})^{1/2} \right\}^{1/3} \quad (4.5)$$

4.3.5 Circular Dichroism (CD)

Proteins were dialysed to Na_2HPO_4 and 50 mM NaCl at pH 7.4, and all CD was done with an Aviv CD spectrophotometer and a 1 mm path-length, quartz cuvette. Constant temperature experiments were done at 25º C, and spectra were recorded in 1 nm steps with a 1 nm bandwidth and 5 seconds of averaging. A series of Python scripts were written to deconvolve individual data sets. The scripts subtracted buffer signal, removed data with a high dynode voltage (> 600.5), and converted data to units of molar residue ellipticity (for scripts see appendix). The molar residue ellipticity of protein mixtures was calculated using:

$$\frac{\text{millidegrees observed in mix} - \text{protein A alone} - \text{protein B alone} - \text{buffer}}{[\text{Protein A}] * (\text{length of A} - 1) + [\text{Protein B}] * (\text{length of B} - 1)} \quad (4.6)$$

where proteins “A” and “B” are arbitrary names, square brackets “[ ]” indicate concentration, and the protein length is in amino acid residues.

4.3.6 Single Molecule Force Spectroscopy

All single molecule constructs were labeled with biotin and an oligonucleotide-CoA modification by methods previously published [189, 252, 290, 452]. In brief, oligonucleotides from IDT were annealed on a thermocycler (70ºC for 10 minutes, 95º C for 2 minutes, then down to 4º C in 0.1º per second steps). The double

stranded oligo-CoA and biotin were then covalently linked to the my protein constructs by Sfp and BirA, respectively (15 µM of target protein and oligo, 30 µM of each enzyme). Labeling was checked by SDS-PAGE. The day before each experiment, oligo-labeled protein was ligated to complementary ∼2.8 µm polystyrene beads, overnight at 15º C, forming a 50 base pair DNA "handle." The opposing polystyrene beads were ligated to 1.8 kilo-base pair DNA that has a biotin-streptavidin on the aqueous end, which binds the biotin of the protein construct. During single molecule experiments, the microfluidics chamber was filled with either buffer alone (100 mM Tris base, 80 mM NaCl, 50 mM arginine HCl at pH 7.4) or buffer plus TSG101cc (a.a. 229—304).

All experiments were of the constant velocity variety. Force was increased from ∼2 pN to 40 or as much as 60 pN to confirm single molecule character (DNA overstretch). Multiple refolding times were tested, from 0 to 30 seconds, and multiple concentrations of TSG101cc were tested, from 1 to 20.5 µM.

4.3.7 Cell Culture

U2-osteosarcoma cells (U2OS) were maintained in McCoy's 5A medium, supplemented with 10% fetal bovine serum and antibiotics (streptomycin and penicillin). Cells were grown at 37º C in a 5% CO_2 atmosphere and the medium was changed every 2–3 days until passaging. Trypsin with EDTA was used for detaching cells during passaging. For transfection, cells were plated at a density of 30000 cells

per well on 96-well plates or 300000 cells per well on 6-well plates and grown to ∼90% confluence before transfecting. Xtreme GENE HP (Roche) was used to transfect DNA per the manufacturer directions.

4.3.7.1 Luciferase Assays

Most luciferase assays were done using 96-well plates, except comparison of TSG101 mutants proceeded by 6-well plates. The NEB *Gaussia* and *Cypridina* plasmids were used at 40 ng each per well for each 96-well assay (or ten times as much for 6-well plates). Two GR response elements were cloned into the promoter of the *Gaussia* plasmid and the *Cypridina* plasmid served as a control [243]. GR expression plasmids were transfected at 0.6 ng per well and TSG101 plasmids were transfected from 0 up to 9.4 ng per well, with the empty pJ603 plasmid balancing the total transfection to 10 ng of pJ603 DNA (96-well plates).

4.3.7.2 Confocal Microscopy

Cells were passaged onto German cover slips in 6-well plates. For the immunostaining data, there were two transfection procedures, both using 300000 cells per well. The first addressed whether or not TSG101 is in the nucleus of U2OS cells, with or without GR. Each pJ603 plasmid (TSG101 and/or GR) was transfected at 100 ng and sonicated salmon sperm DNA was used to bring the total amount of DNA to 1000 ng per well, and the cells were fixed about 40 hours

after transfection. The second procedure addressed how TSG101 localization changes in response to the luciferase assay conditions, over time. Transfection proceeded in the exact same manner as with luciferase assays, except with a total of 500 ng of DNA per well (10x that used on 96-well plates). Cells were then collected for fixation 16 or 40 hours later.

TSG101-tdTom was used as an alternative verification that TSG101 is in U2OS cell nuclei. For expression of TSG101-tdTom, 100000 cells were passaged and the expression plasmid was transfected at 15 ng per well and sonicated salmon sperm DNA was used to bring the total DNA amount per well to 500 ng. The TSG101-tdTom cells were fixed $\sim$24 hours after transfection.

All immunostaining was done at room temperature. Cells were washed thrice with PBS+2 mM $MgCl_2$ (PBSM), fixed with 4% PFA for 10 minutes, and washed thrice again with PBSM. The PFA was quenched for 10 minutes using 50 mM NH_4Cl in PBSM. Parafilm was stretched across the bench, and the cover slips were set cell-side up on top of the parafilm. The top of the cell culture plate was used to cover the slides and make a humid chamber.

The cells were permeabilized and DNA was first stained for 30 minutes with 150 µL of: PBS, 0.1% TX-100, 1% BSA, 0.5 µg/mL DAPI. Then the cover slips were incubated for one hour with primary antibody at a 1:500 dilution (GR primary was rabbit mAb from Cell Signaling D8H2; TSG101 primary was mouse mAb from GeneTex 4A10). The cover slips were washed thrice for five minutes with

PBSM and incubated with secondary antibody at a 1:600 dilution for 30 minutes in the dark. The GR secondary was Life Technologies A11034 (goat anti-rabbit), labeled with Alexa Fluor-488, and the TSG101 secondary was Cell Signaling 8890S (goat anti-mouse), labeled with Alexa Fluor-594. Cover slips were then washed thrice for five minutes with PBSM. All antibody dilutions were made using the aforementioned DAPI solution. Cells expressing TSG101-tdTom were DAPI stained as above, without the antibody steps. The cover slips were mounted using flouromount (Sigma).

All microscopy was done using the Integrated Imaging Center of Johns Hopkins University. The data shown were collected using either a Zeiss LSM 700 (immunostaining) or 780 (TSG101-tdTom) confocal microscope. DAPI was excited with a 405 nm diode laser, AlexaFluor 488 (GR) was excited with a 488 nm laser, and both AlexaFluor 594 and tdTom (TSG101) were excited with a 561 nm laser. All imaging was done at either 400x or 630x magnification. The gain and digital offset were optimized for the best signal-to-noise in each image; thus, the absolute intensities are not necessarily comparable. Analysis of images was done using the Fiji version of ImageJ [356, 357].

4.3.7.3 W-blots

Transfections were similar to the luciferase assays, except always on 6-well plates. Cells were collected by removing medium, washing with PBS, and scraping cells

off the bottom. After pelleting the cells by centrifugation (194 rcf, one minute), the cells were lysed using Cell Signaling Technology's lysis buffer supplemented with 350 mM NaCl. A 26 gauge needle was used to break apart the cells by rapid plunging. The lysate was cleared by centrifugation at 16873 rcf for 30 minutes. Total protein content of the supernatant was measured by Bradford assay, with BSA as a standard (BioRad), and this was used to normalize the amount of total protein loaded in each lane of a 4–15% Mini-PROTEAN TGX precast gel (BioRad). A fluorescent ladder was used to detect transfer efficiency and estimate molecular weights (Spectra multi-color, Thermo).

After running SDS-PAGE to completion, gels were transferred to methanol-activated PVDF membranes (low-autofluorescence, BioRad) at 100 V for 25 minutes on ice. Blots were then cut horizontally along the 95 kDa ladder mark so P150 (loading control) could be blotted separately from TSG101 and GR. All blocking was done at room temperature for 30 minutes, using ∼4% non-fat dry milk in PBS that was cleared for five minutes by centrifugation.Blots were then washed thrice for 5 minutes with PBS+0.1% Tween-20 (PBST) before incubation with primary antibodies. The primary and secondary antibodies for GR and TSG101 were the same as the ones used for microscopy. The P150 antibody is a mouse mAb from BD Biosciences (#610473) and the TSG101 secondary was used to label P150 (anti-mouse Cell Signaling 8890S). Primary antibody incubations were overnight at 4º C in PBST with 1–4% milk (no difference observed). Primary

antibody dilutions were 1:5000 (P150), 1:2000 (TSG101), and 1:2000–4000 (GR, no difference observed). Blots were washed thrice for 5 minutes with PBST before incubating at room temperature with secondary antibodies in PBST with 1–4% milk (1:5000 P150, 1:5000–6000 GR, 1:2000–5000 TSG101; TSG101 was difficult to perceive by naked eye with a 5000 fold dilution, but was detectable). Blots were washed thrice for five minutes with PBST and excess liquid was gently drained before letting the blots dry in the dark.

Imaging of the fluorescent blots was acheived with a FluorChem from Protein Simple (generously shared by Prof. Wendland at Johns Hopkins). To analyze the images, I first normalized signal in each lane to its loading control, then normalized signal of each protein to its signal with 0 ng of TSG101 transfected. In this way, blot-to-blot variations in absolute signal were minimized and I compared only relative changes.

4.4 Results

4.4.1 TSG101cc Binds the NTD of GR

4.4.1.1 Protease Protection

A yeast two-hybrid study indicated that the coiled-coil of TSG101 (TSG101cc) bound a region of GR's disordered N-terminus, called the Activation Function-1

(AF1), close to the C3-NTD start site [153]. I first sought to confirm the binding of these proteins *in vitro* and in the process also verify the proposed binding site on GR. Disordered proteins, such as the GR NTD, will rapidly degrade in the presence of trypsin; however, binding partners can protect disordered proteins from degradation [219, 338]. The mechanism appears to be coupled binding and folding, as discussed later in this chapter and by others [216, 217, 219, 242, 338].

In Figure 4.3 I digested either the A- or C3-NTD of GR in the presence or absence of TSG101cc (see controls in Figure 4.4). By mass spectroscopy, I was able to reproducibly observe fragments left by C3-NTD, and this is shown in part B of Figure 4.3 (raw data in Figure 4.5). Note that the protected fragment observed here, matches the bait used by others in a yeast two-hybrid [153]. MALDI of the A-NTD fragments was attempted multiple times, but did not yield a strong signal; nonetheless, the gels indicate that the same region of the NTD seems to be protected by TSG101cc in both the A- and C3-NTD. In both cases, the presence of TSG101cc protected a series of fragments between about 14.8 to 20.9 kDa in size. This contrasts with the 9.1 kDa long fragment I actually observed by mass spectroscopy of the protected bands (Figure 4.3B), and I am likely missing some ions in this experiment (see cross-linking below).

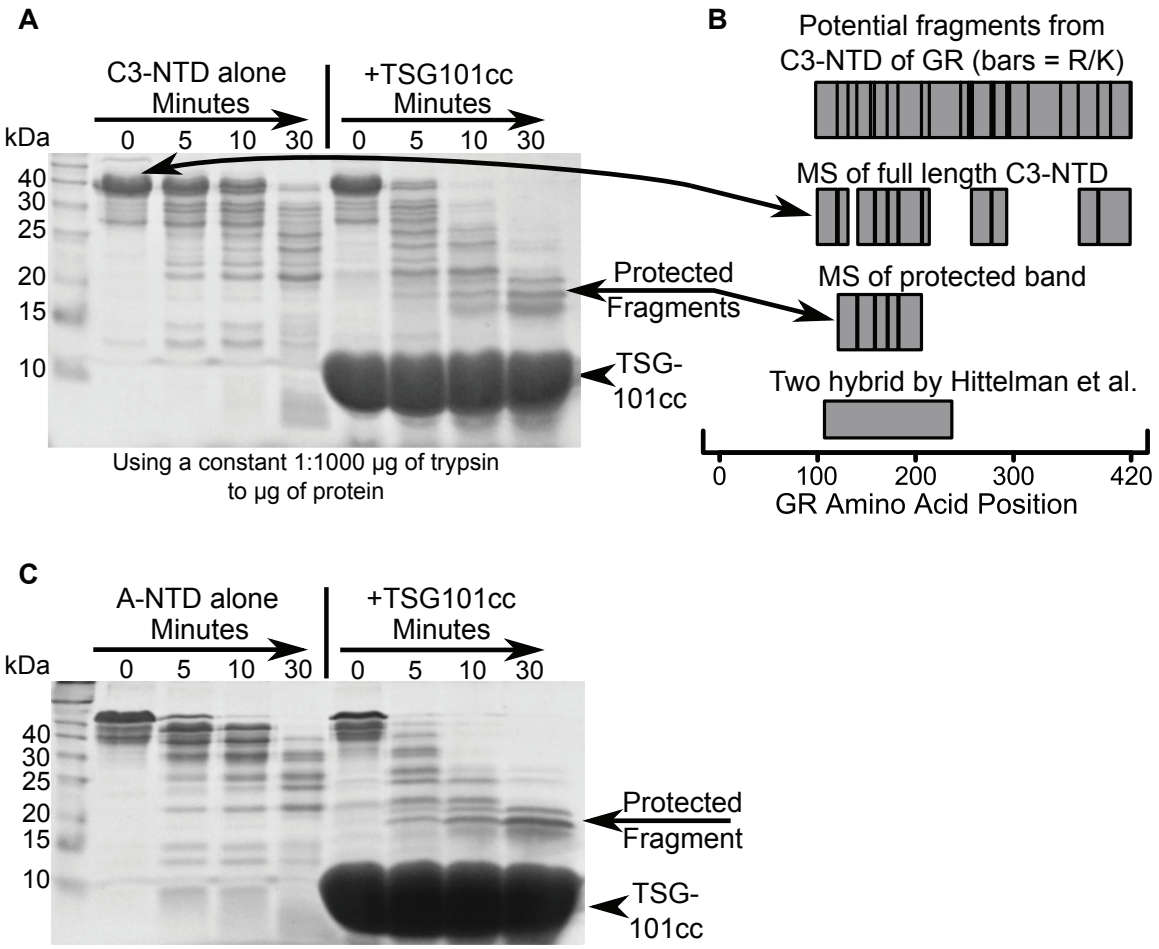


Figure 4.3 – TSG101cc Protects the N-terminus of GR from Protease Digestion
Both gels shown here were coomassie stained. **A)** Digestion of the C3-NTD alone (left) produces no specific fragments. Addition of TSG101cc and more trypsin (right side, using a constant trypsin:target ratio) causes formation of a protected fragment. **B)** Results of MALDI mass spectrometry of bands cut from a gel, as shown in part A, and fully trypsinized overnight. Possible fragments are shown at the top with trypsin cut sites (Arg or Lys) as black bars. The next two sets of boxes represent the fragments actually observed for full-length C3-NTD and the protected band, respectively. The box at the bottom represents the sequence used as bait in the yeast two-hybrid by Hittelman et al. [153]. **C)** A trypsin digestion gel, as in part A, except using the A-NTD of GR.

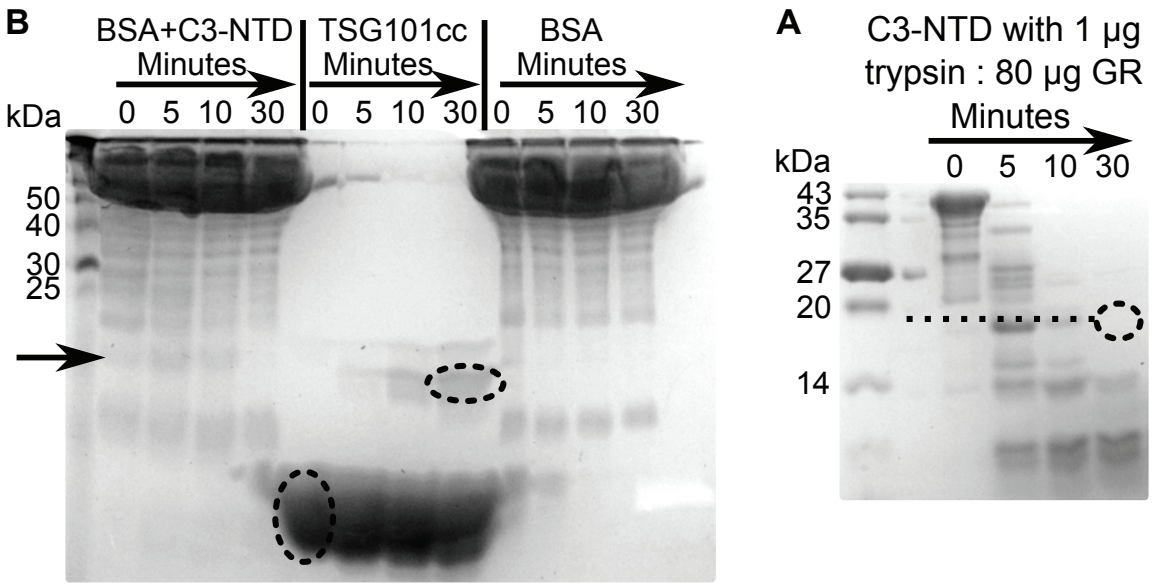


Figure 4.4 – **Control Trypsin Digests**
Coomassie stained gels are shown: **A)** Protein was digested as in Figure 4.3. In the BSA control, the arrow indicates a possible GR fragment that disappears by 30 minutes (compare to BSA alone). The dashed circles indicate bands that were cut out and fully trypsinized for MALDI shown in Figure 4.5 parts A and B. Note the formation of an upshifted TSG101cc digestion product (see Figure 4.5B). **B)** C3-NTD was digested alone with the same ratio of trypsin that was used in the presence of TSG101cc in part A of Figure 4.3. The dashed line and circle indicate where the protected fragment was previously (∼19 kDa).

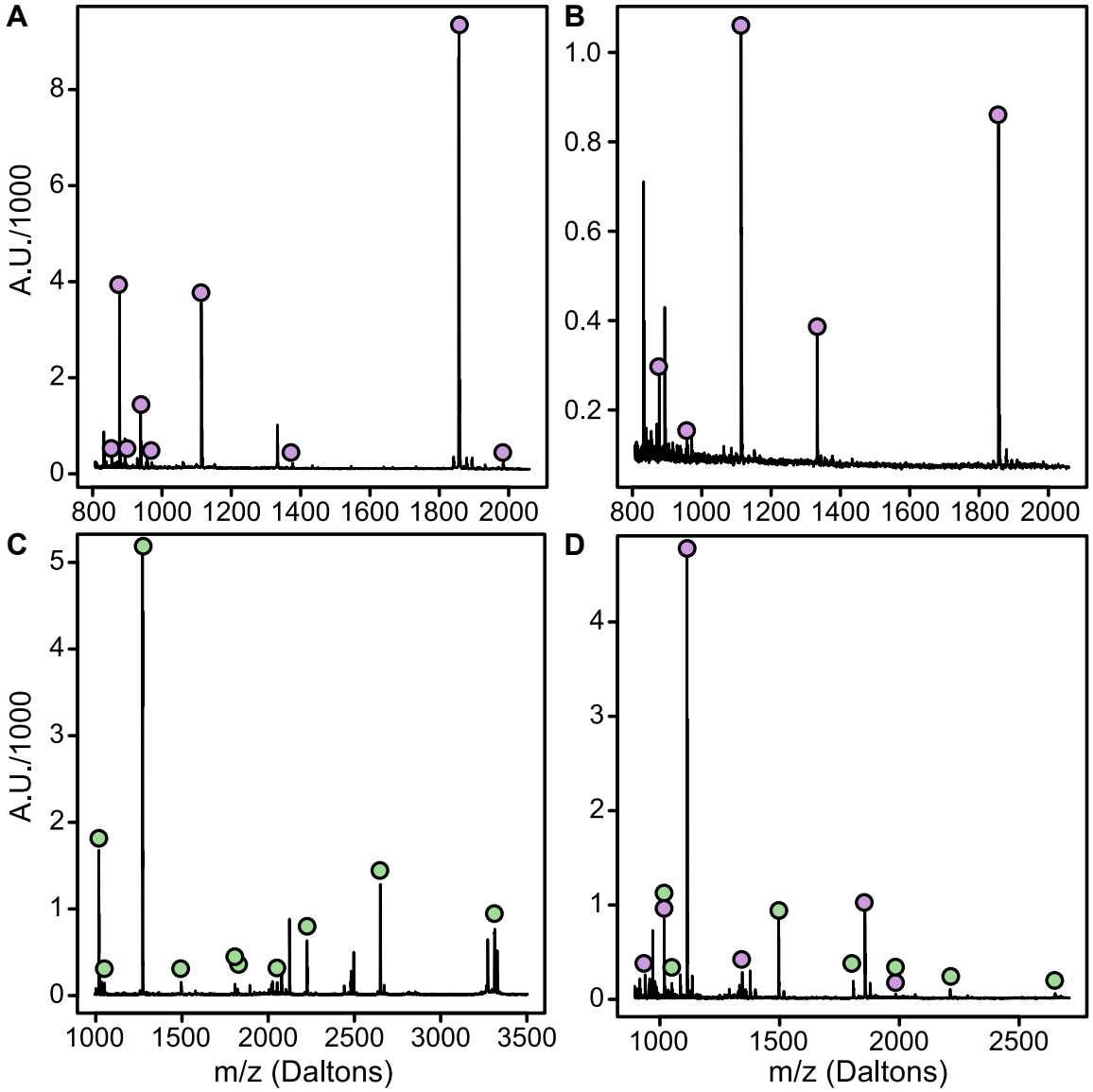


Figure 4.5 – **Mass Spectrometry of Protease Protection Assays**
X and Y axes are the same units for all the plots. Ions are labeled with green circles for those consistent with C3-NTD and purple circles for those consistent with TSG101cc. Note that some peaks in part D are ambiguous and were excluded from Figure 4.3B. **A)** The undigested TSG101cc band, running at ∼9 kDa in SDS-PAGE (see Figure 4.4). **B)** The digested TSG101cc band running close to 19 kDa produces ions consistent with at least a.a. 1—66 (also see Figure 4.4). **C)** The undigested C3-NTD band (Figure 4.3A). **D)** The band protected in the C3-NTD + TSG101cc experiment (Figure 4.3).

4.4.1.2 Cross-linking

Cross-linking of proteins can reveal the overall structural organization or even a low-resolution structure [18, 454]. Using cross-links as constraints in *ROSETTA* has even been used to aid *de novo* structure prediction [146, 188]. By using cross-linkers with multiple lengths, one can build a robust list of constraints. Here I tested three cross-linkers for TSG101cc and the C3-NTD, but BS3 was overly effective at cross-linking and EDC yielded no cross-links (11.4 Å and zero-length cross-linkers, data not shown). Under carefully constrained conditions, DST produced a single TSG101cc–C3-NTD cross-link without a large amount of background (Figure 4.6). In support of the band being an oligomer of TSG101 and GR, ions indicative of both species were present in mass spectrometry (Figure 4.6B). The data also contain several cross-link ions, shown in Figure 4.6B and in cartoon schematic in Figure 4.7. In the previous section I showed that TSG101cc protects a fragment of C3-NTD that may be as large as 20.9 kDa by SDS-PAGE. Combining the mass spectrometry results of the protease protection and the cross-linking studies shown here yields a ∼19 kDa stretch of the C3-NTD.

Based on these observations, I propose that GR and TSG101 form a pseudo-trimer, namely the GR NTD wraps back on the TSG101cc in a manner reminiscent of the ESCRT-I coiled-coil [212] (Figure 4.7). This would likely satisfy the hydrophobicity of the TSG101cc, a requirement for a stable oligomer of TSG101cc

as I showed previously [438]. The exact alignment of the binding is unclear, but it would likely leave the TATA-binding protein (TBP) site on GR open to bind TBP and TSG101's N-terminus would likely face the ubiquitination site of GR (close to the DBD [424]). Others have shown that TSG101 can stabilize the monoubiquitinated forms of androgen receptor and the E3 ligase MDM2 [63, 245]. My tentative model would allow for a similar interaction in GR.

4.4.2 Binding of the GR NTD and TSG101cc is Coupled to Conformational Change

Several studies have shown that when transcriptional cofactors bind the disordered GR NTD, they do not simply occlude the NTD from solvent and protect it from proteases. Instead, transcriptional cofactors bind and fold the disordered NTD of GR [201, 216, 219], and this seems to be a general property of cofactor binding to the NTD's of many steroid hormone receptors [217, 430]. Here I tested whether TSG101cc can promote the folding of the GR-NTD. I used circular dichroism (CD) and single molecule force spectroscopy (SMFS, colloquially called optical tweezers). Under ideal conditions these techniques yield quantitative data [201, 252]. Due to experimental circumstances, my data are of a more qualitative variety, but nonetheless support folding of GR when it binds TSG101cc.

I first used CD to measure the folding of GR. It must be noted that CD will

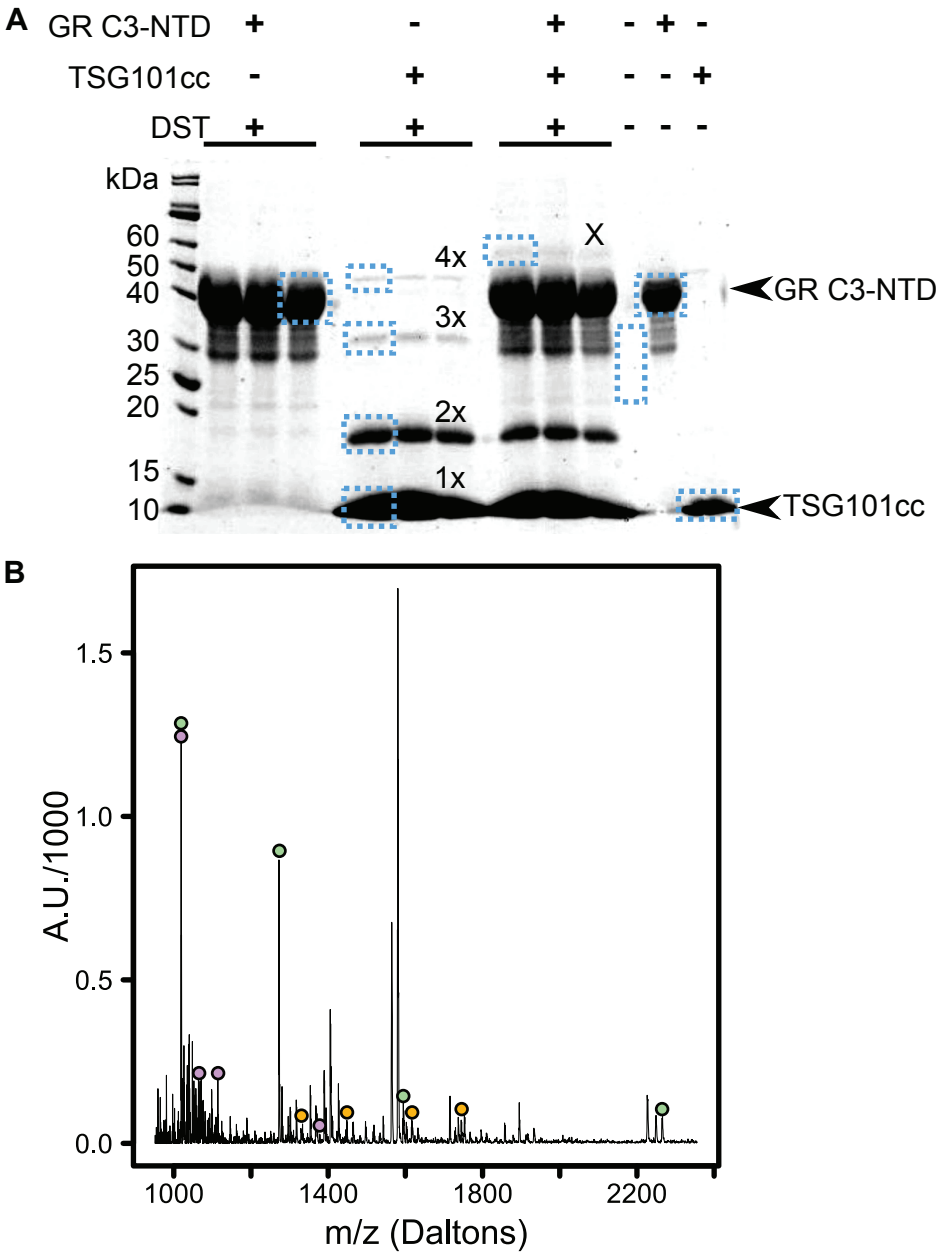


Figure 4.6 – **Cross-Linking of C3-NTD and TSG101cc**
A) An example SDS-PAGE of my DST cross-linking reactions is shown with coomassie staining. The GRxTSG101 cross-link was barely visible by eye and the contrast of the whole image has been heavily shifted to make it visible in the image for the reader. The DST samples were loaded in triplicate. TSG101cc cross-links with itself, forming the homotetramer I recently reported [438] (1x, 2x, 3x, 4x are the oligomeric cross-links). The C3-NTD—TSG101cc cross-link has a "X" directly above it. Because the gel is overloaded, it is difficult to tell the molecular weight of the cross-link, but it appears to be ∼9 kDa larger than C3-NTD—consistent with a C3-NTDxTSG101cc heterodimer. Blue boxes indicate the bands/controls cut out for MALDI. **B)** An example MALDI dataset of a cross-linked band that was treated with both trypsin and chymotrypsin. Ions are labeled with colored circles as follows: green = C3-NTD, purple = TSG101cc, orange = cross-link.

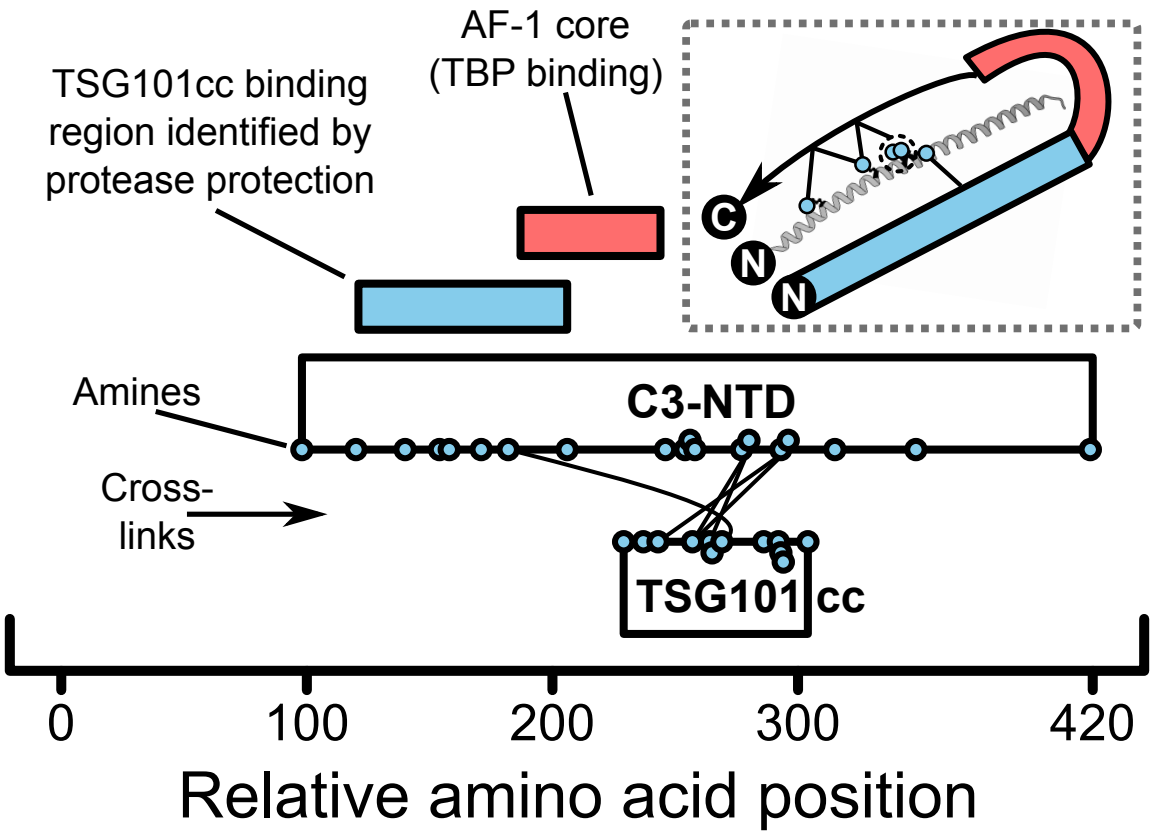


Figure 4.7 – Cross-Linking Results and Speculative Model
All boxes in the main figure are scaled and positioned relative to the respective protein's full-length sequence. The white boxes represent the GR and TSG101 constructs used for cross-linking and the lines between them are the cross-links discovered. The cross-linker is amine reactive and amines are shown as blue dots (lysines or N-termini). The blue box is in reference to Figure 4.3 and the red box at the top is in reference to the literature [11]. **Inset:** A tentative model of the structural organization of the heterodimer, based on the evidence presented here. Colored boxes are as in the main figure, but note that the protected region (blue, 86 a.a.) is slightly longer than the TSG101cc (76 a.a.). The TSG101 coil is in gray and the cross-linked lysines are shown as blue dots with lines going roughly to the appropriate spot on GR (alignment unclear). One cross-link was ambiguous because of adjacent lysines—it is indicated by a dashed circle. The black spheres "N" and "C" indicate N and C-termini. Note that the TSG101 N-termius is pointed in the direction of the rest of the GR protein. This allows for the possibility that TSG101's UEV domain could bind the ubiquitin that GR receives close to the DBD [424].

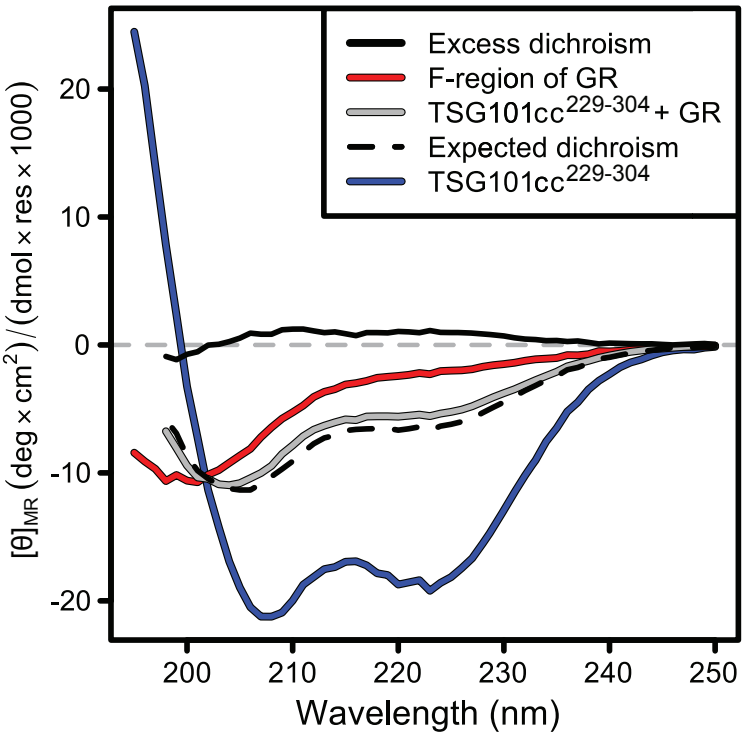


Figure 4.8 – CD Reveals a Structural Change in Binding of TSG101cc and the C3-NTD
A small loss of helicity is observed when mixing 6 μM C3-NTD with 10 μM TSG101cc. This suggests that upon binding GR, this construct of TSG101cc does not regain all the helicity present in the TSG101cc homotetramer. This is nonetheless indicative of binding.

actually report on structural changes in both GR and TSG101, and only by saturating the system with one or the other can the actual change in structure be extracted. Saturating the GR NTD with TSG101cc was not possible in CD without saturating the dynode voltage; thus, the CD data are approximations of the structural changes. I was able to strengthen the signal from CD by using a longer TSG101cc construct that was shown to bind GR in a yeast two-hybrid [153]. Please note that this is the only section that makes use of this longer TSG101cc construct, TSG101cc$^{229\text{-}339}$ (a.a. 229—339). Otherwise, the rest of this chapter uses the shorter construct because of higher yields from *E. coli*.

Both TSG101cc constructs produced structural changes in the presence of C3-NTD (and Figure 4.9). The TSG101cc$^{229\text{-}304}$ construct appears to lose a small amount of helicity upon binding GR Figure 4.8. This suggests that the GR-bound complex contains slightly less structure than the TSG101cc homotetramer. In contrast, TSG101cc$^{229\text{-}339}$ produced a large increase in helicity upon mixing with the C3-NTD (perhaps $\sim$36 a.a. of α-helix [132]), but not with my negative control, the AF1 core that binds TBP. The TSG101cc$^{229\text{-}339}$ extends partly into the Steadiness domain of TSG101 and is less ordered than the TSG101cc$^{229\text{-}304}$ construct. It is possible that binding of TSG101cc$^{229\text{-}339}$ and C3-NTD folds both proteins to some degree.

To directly measure folding of the GR NTD, I used SMFS. As shown in Figure 4.10, in the presence of TSG101cc, I was able to observe unfolding transitions

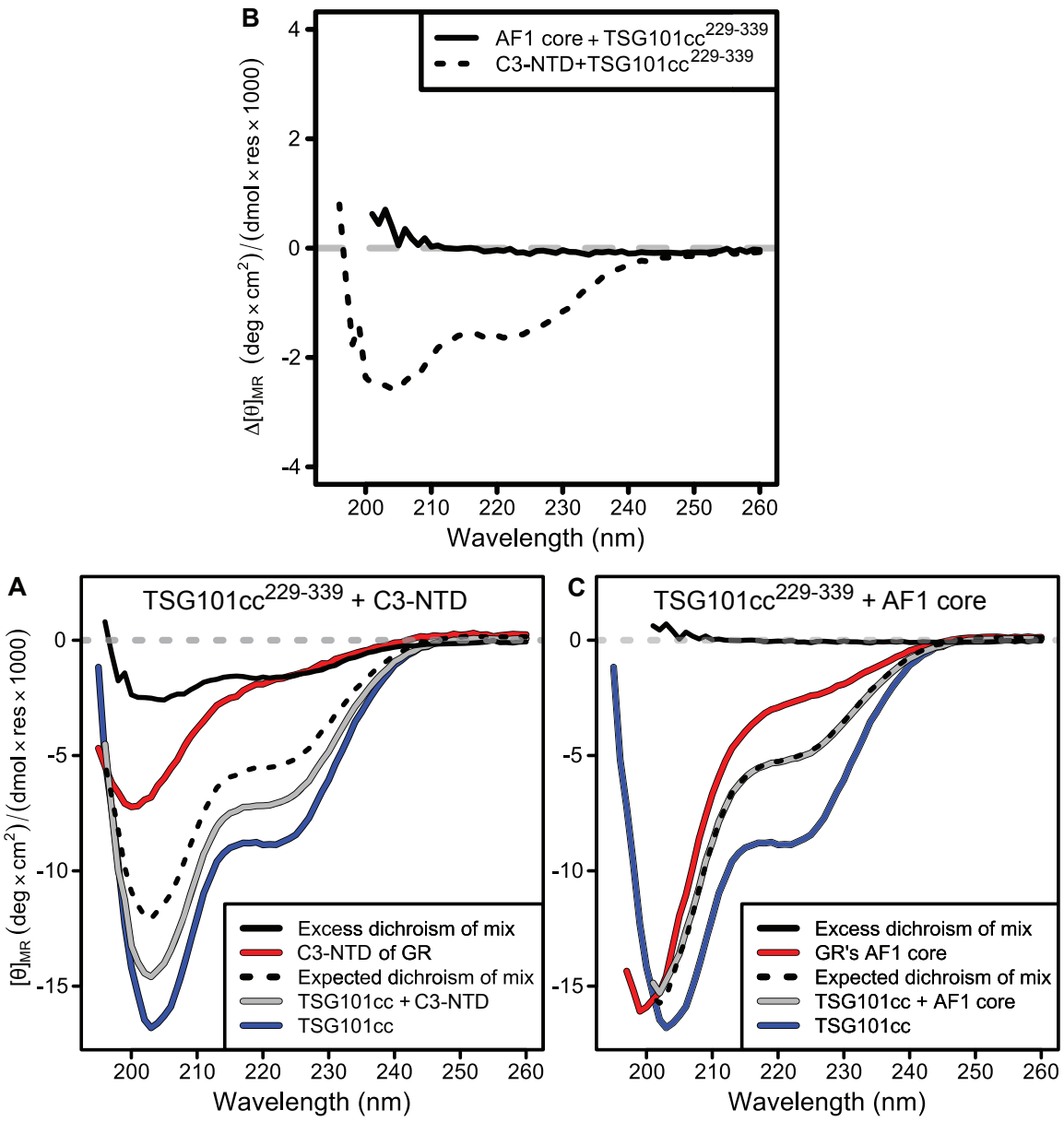


Figure 4.9 – CD Reveals a Structural Change in Binding of TSG101cc and the C3-NTD
The data here used a longer TSG101cc construct (a.a. 229—339) than elsewhere in this chapter. It is the same as was used in a yeast two-hybrid of TSG101 and GR [153]. **A)** Mixing the TSG101cc$^{229\text{-}339}$ with the C3-NTD of GR, but not the AF1 core, produces an excess dichroism (C3-NTD = 3 μM, TSG101cc$_{229\text{-}339}$ = 20 μM, AF1 core = 58.4 μM shown here). See **B)** and **C)** for the raw data. The data in A are replotted as solid black lines in B and C for comparison. The gain in helicity seen in part A is equivalent to ∼36 a.a. folding into α-helix [132].

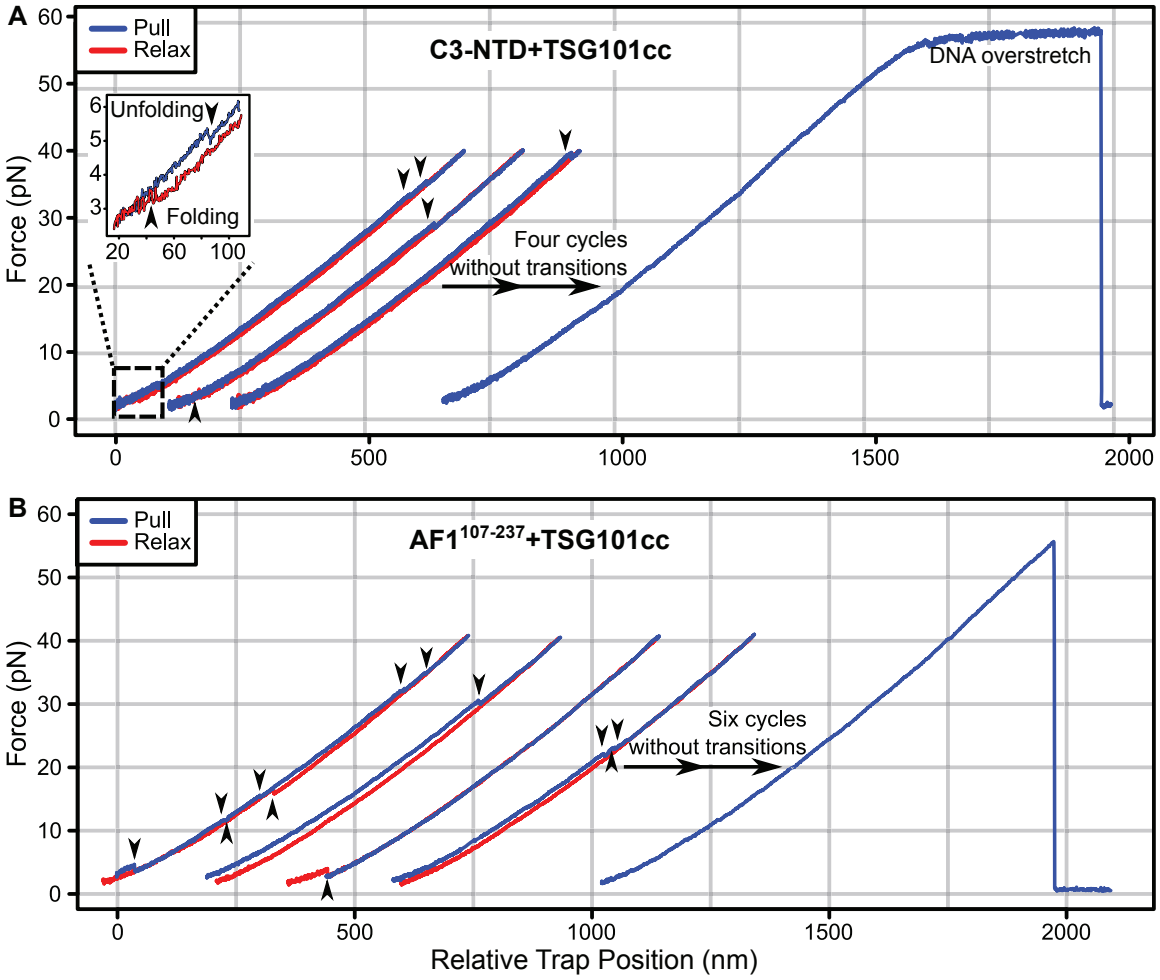


Figure 4.10 – **TSG101cc can Fold the N-terminus of GR**
The data are offset on the x-axis for visual purposes. Unfolding and folding transitions are highlighted with arrowheads facing down or up, respectively. **A)** Force ramp data of the C3-NTD in the presence of 10 μM TSG101cc. The inset graphic is a zoom in of the low force regime for the first trace. **B)** Force ramp data of $AF1^{107-237}$ in the presence of 10 μM TSG101cc.

of both the C3-NTD and the AF1 construct used in a yeast two-hybrid of TSG101 and GR, $AF1^{107\text{-}237}$ [153]. Without TSG101cc present, I observed only one transition of a confirmed single molecule of GR (Table 4.2). It must be noted that SMFS can produce artifacts from non-specific adsorptions (Prof. C. Kaiser communication). However, I observed no transitions of my AF1 core construct, with or without TSG101cc (Figure 4.11). This suggests that TSG101cc alone cannot create the transitions I observed and supports my hypothesis that these transitions are largely unfolding events of my GR NTD constructs.

Experiment	Total # of single molecules	Total # of pull cycles	# of single molecules with transitions	Total # of transitions
AF1 core	8	46	0	0
AF1 core + TSG101cc	9	84	0	0
C3-NTD	8	33	1 (13%)	1 (3% of pulls)
C3-NTD + TSG101cc	28	246	13 (46%)	63 (25%)
$AF1^{107\text{-}237}$	5	40	0	0
$AF1^{107\text{-}237}$ +TSG101cc	10	83	3 (33%)	16 (13%)

Table 4.2 – Compilation of SMFS results
Addition of TSG101cc to the buffer produced transitions in the C3-NTD and $AF1^{107\text{-}237}$.

Normally, when collecting SMFS data, one wishes to collect thousands of transitions in order to access quantitative analyses [98, 266]. Because GR's NTD is normally unfolded, my default measurement state is essentially void of transitions. My hypothesis is that only by random chance of TSG101cc binding

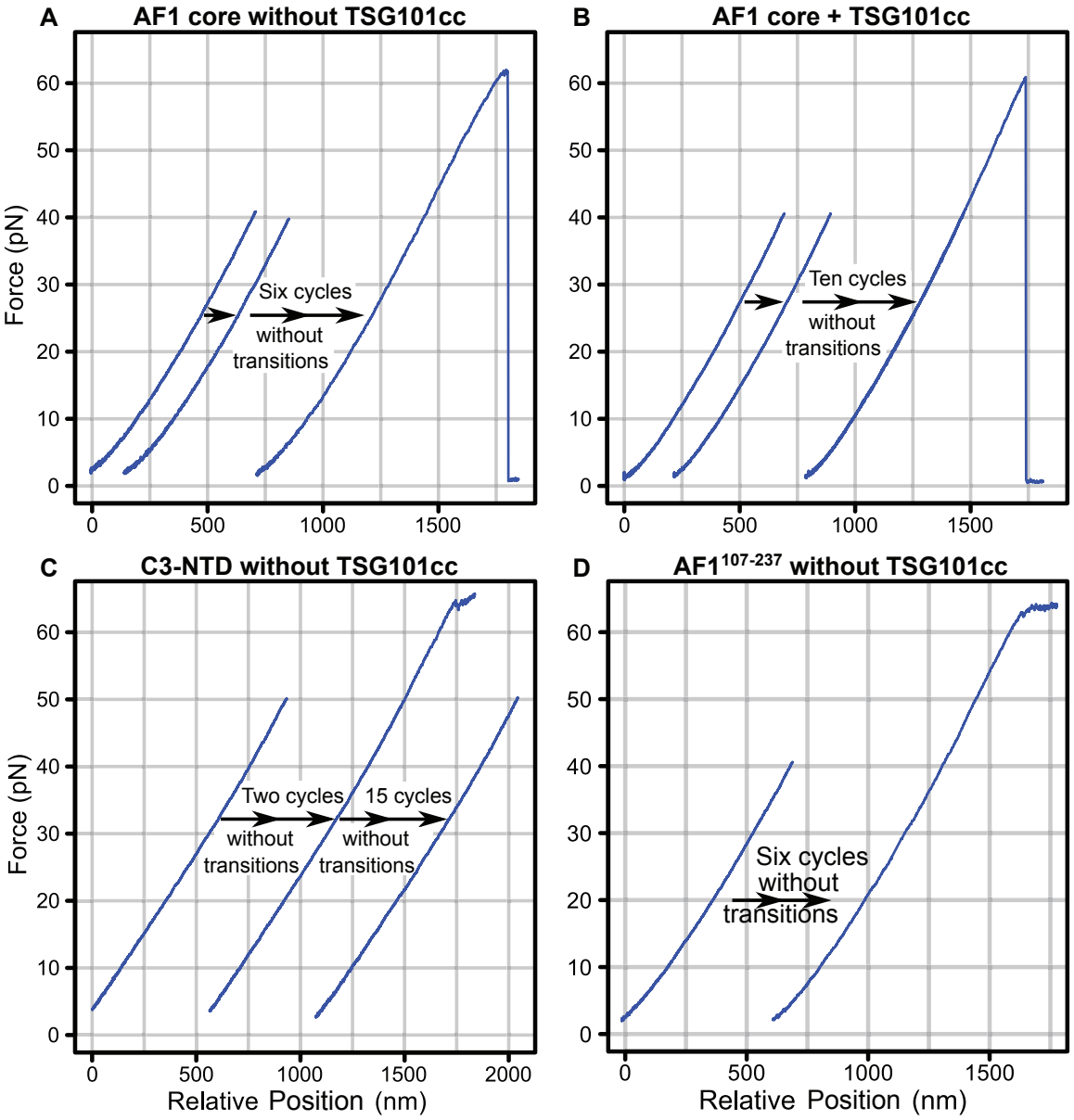


Figure 4.11 – **Transitions Seen in SMFS Depend Upon the TSG101cc Binding Site of GR** Representative data are presented with individual traces offset on the x-axis for visual purposes. **A)** and **B)** are data for the AF1 core (binds TBP, not TSG101cc) without or with 10 μM TSG101cc. **C)** and D) are control data for C3-NTD and the AF1$^{107-237}$, respectively.

GR, do I see transitions. This has resulted in a much lower experimental yield than usual (Table 4.2), and as such I cannot make quantitative statements about these folding events. Figure 4.23 in the Appendix provides a graph of my observed contour lengths, which were all less than the length of the C3-NTD.

4.4.3 TSG101cc Allosterically Promotes DNA-Binding of GR

Our previous work showed that the C3 isoform is more stable than the A isoform because of the destabilizing R-region in the A isoform [242]. Because it would cost less energy to fold the C3 NTD, I predicted that transcriptional cofactors, namely TSG101cc, would bind C3 better than A. To measure the binding of GR and TSG101cc I labeled the TSG101cc with pyrene, as described before [438], and observed fluoresence anisotropy of pyrene-TSG101cc as I titrated an NTD isoform out of solution. As seen in Figure 4.12A, the A and C3 isoforms of GR bind TSG101cc equally well with a K_d of about 20 μM (see Table 4.3 for the fitted values). Though my hypothesis proved false for TSG101 (see Discussion), the two proteins do bind and I decided to proceed with other tests of this interaction.

Numerous labs working with different steroid hormone receptors have shown that binding DNA causes a conformational change in these proteins, often stabilizing their N-terminal domains [28, 29, 58, 131, 167, 215]. Thermodynamics is a two-way street and I hypothesized that if binding DNA can stabilize a structured state of the NTD, then the NTD binding TSG101cc should promote DNA-

binding [75, 403]. If not, then TSG101 is unlikely to be directly involved in the transcriptional functions of GR, and may only have an indirect effect through regulation of GR degradation and/or post-translational modification [299]. As seen in Figure 4.12B and Table 4.3, binding of TSG101cc promotes full-length GRE-binding of both the A and C3 isoforms of GR. This is a decidedly allosteric effect, as TSG101cc itself does not bind DNA (see the unbound baselines of the raw data in Appendix, Figure 4.25)

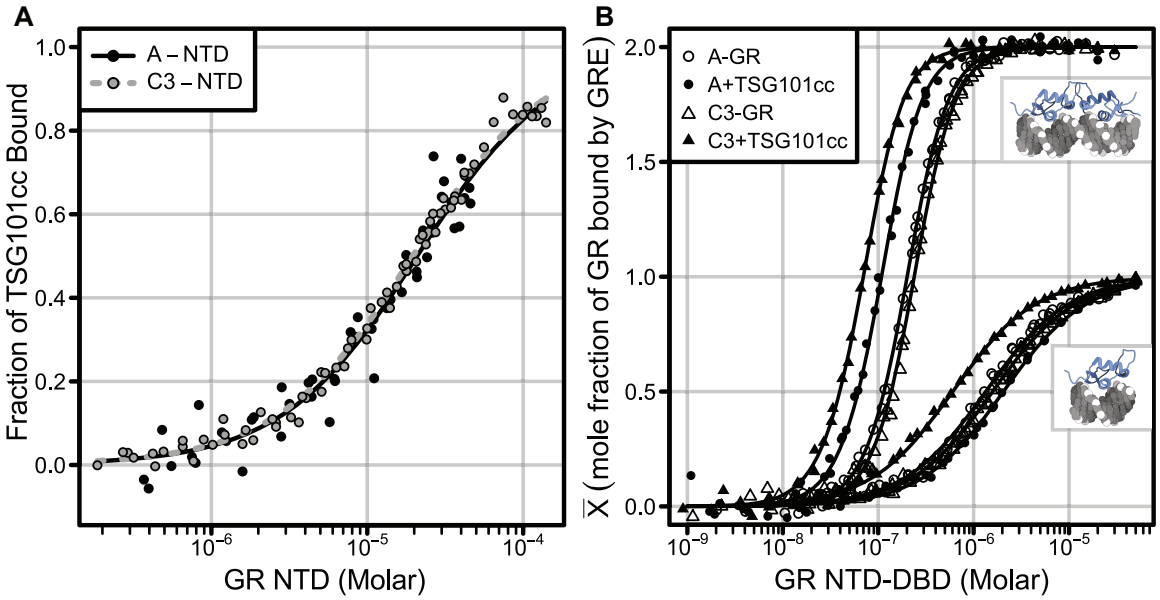


Figure 4.12 – Quantitative Binding of TSG101cc to GR
All of the data shown here are from fluorescence anisotropy that was globally fit and then converted to mole fraction bound. See Table 4.3. **A)** Fluorescence anisotropy of pyrene labeled TSG101cc was observed while titrating one of the NTD isoforms listed. The data were fit to a single-site binding model. **B)** Fluorescence anisotropy of 6FAM-labeled GRE or GRE half-site binding to the listed NTD-DBD isoforms of GR, in the presence or absence of 100 μM TSG101cc. The DNA element used for each data set is indicated by the cartoon of GR DBD (blue) binding DNA as a dimer or monomer [263]. The GRE half-site data were fit to a single site binding model. The full GRE data could only be fit using an all-or-none, two-site binding model.

TSG101cc stabilizing GR's binding of DNA is itself interesting, but the full-length GRE hides the underlying mechanism of the allostery. To give mechanistic

GR	TSG101cc	6FAM-GRE	K_d
A-NTD	+ pyrene	N/A	20.9 ± 4.7 μM (3)
C3-NTD	+ pyrene	N/A	19.6 ± 1.0 μM (3)
A-NTD-DBD	–	Half	1.33 ± 0.12 μM (3)
A-NTD-DBD	+	Half	2.05 ± 0.37 μM (2)
A-NTD-DBD	–	Full	201.9 ± 8.6 nM (3)
A-NTD-DBD	+	Full	98.3 ± 8.9 nM (3)
C3-NTD-DBD	–	Half	1.58 ± 0.68 μM (4)
C3-NTD-DBD	+	Half	0.599 ± 0.114 μM (2)
C3-NTD-DBD	–	Full	228.5 ± 11.8 nM (3)
C3-NTD-DBD	+	Full	59.9 ± 4.9 nM (2)

Table 4.3 – **Table of Fitted Binding Constants**
The first two rows refer to binding of GR and TSG101cc in Figure 4.12A and the rest of the table refers to binding of GR and DNA in Figure 4.12B. Parentheses next to the fitted K_d's indicate the number of data sets that were globally fit. For the A and C3 isoforms of GR respectively, TSG101cc induces a –419.2 and –779.8 cal/mol shift in binding to the full GRE, and a +252 and –565 cal/mol shift in binding to the half-site GRE.

insight to this work, I next measured the same DNA-binding effect except with the GRE half-site. This is DNA that has only one GR binding site on it, illustrated in the lower right of Figure 4.12B. If TSG101cc stabilizes the dimerization of GR monomers, using a GRE half-site will ablate that possibility, resulting in no change or perhaps a loss of binding with TSG101cc present. Alternatively, TSG101cc may allosterically promote a DNA-binding conformation of the DBD. Given a $\sim$1 kcal/mol shift in binding for the full GRE (C3 NTD-DBD ± TSG101cc), I expected a $\sim$0.5 kcal/mol shift in binding for the half-site GRE. This is indeed what I observed for the C3 NTD-DBD construct, but not the A NTD-DBD construct. Thus, it appears that the A and C3 isoforms use different allosteric mechanisms to govern DNA-binding (see Discussion).

4.4.4 TSG101 Localizes to the Nuclei of U2OS Cells

The Hilser lab uses U2OS cells because they normally do not express GR, thereby giving us a blank slate on which to express just one GR isoform. Unfortunately, the author is unaware of literature describing TSG101 localization in U2OS cells. Before determining how TSG101 affects GR's activity inside live cells, I needed to verify literature reports that TSG101 can localize to the nucleus of several other cell lines [63, 74, 293, 449, 459]. Shown in Figure 4.13 I transfected U2OS cells with plasmids expressing an NTD-DBD isoform of GR and/or the full-length TSG101 (used *in vivo* except where noted). TSG101 localizes to the nuclei

endogenously and when overexpressed in these conditions, regardless of the expression of GR. Similar results were obtained when overexpressing a tdTom tagged version of TSG101 (Figure 4.14).

The above microscopy experiments used transfection methods slightly different from my luciferase assays, below. To control for any effects of the luciferase assay conditions, I also transfected cells with the full luciferase assay complement (*Gaussia* and *Cypridina* luciferases, a GR isoform, TSG101). As shown in Figure 4.15 and Figure 4.16, a cytosolic aberration appears about 40 hours post-transfection in the group with the highest amount of TSG101 plasmid. This coincided with cell death, as seen by floating cells in the associated wells. Overexpression of TSG101 previously caused both of these effects in literature data [127, 459]. My analysis of TSG101's nuclear localization indicates that at high expression levels, TSG101 appears to saturate its nuclear binding sites and spills over to cytosolic sites (Figure 4.17).

I sometimes observed repression of GR's transcriptional activity at longer post-transfection times, possibly because of these artifacts. To avoid these problems my luciferase measurements (below) were taken 40 hours post-transfection or earlier. Because both of my luciferases are constitutively excreted, my luciferase measurements are largely a time average of everything occurring before the shown artifact.

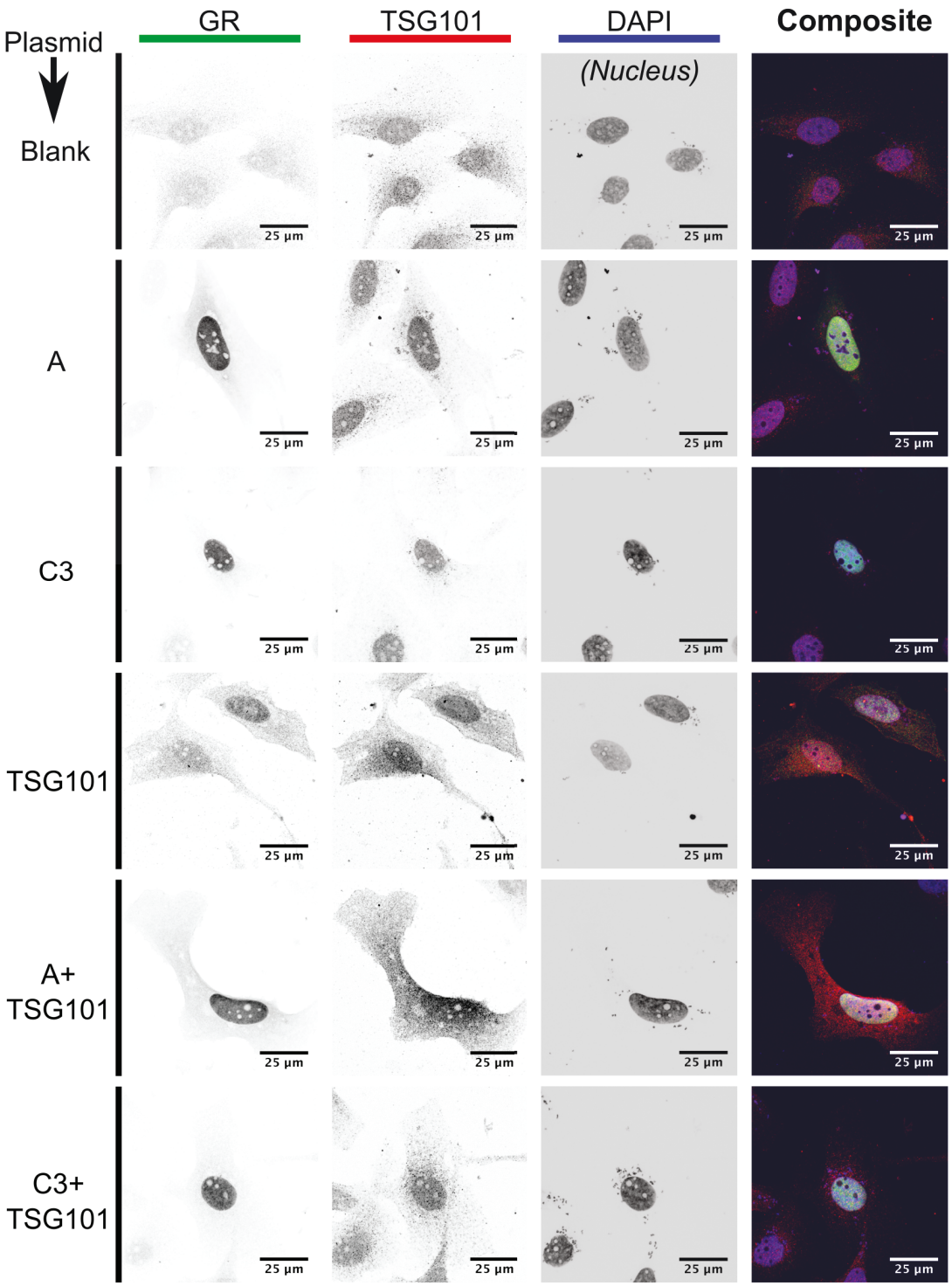


Figure 4.13 – TSG101 Localizes to U2OS Nuclei
The expression plasmids transfected are on the left of each image series. In order to view the GR channel of cells without a GR expression plasmid, the microscope gain had to be increased about 1.4 fold. All scale bars are 25 µm. GR and TSG101 were detected by immunostaining.

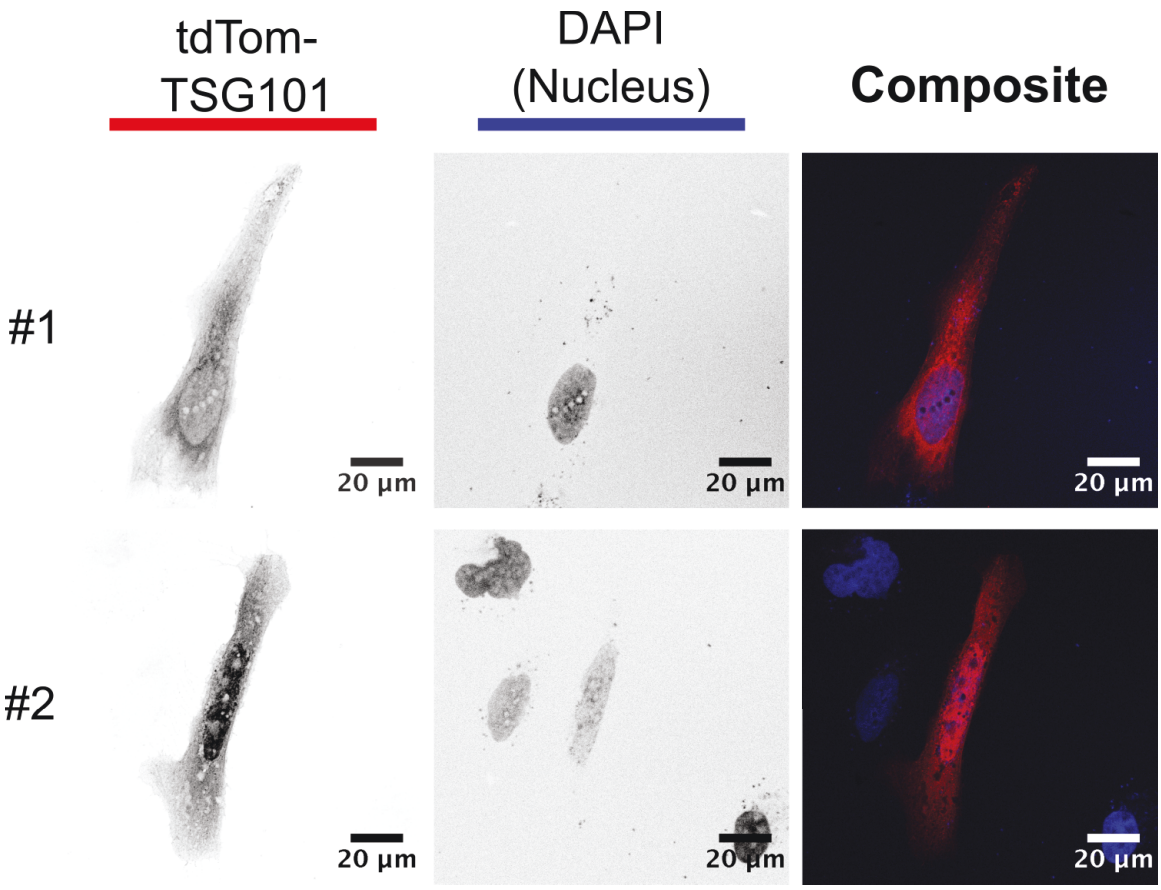


Figure 4.14 – **TSG101-tdTom Localizes to U2OS Nuclei**
Two example image sets are shown to give the reader a sense of the variability in the data set. All scale bars are 20 µm.

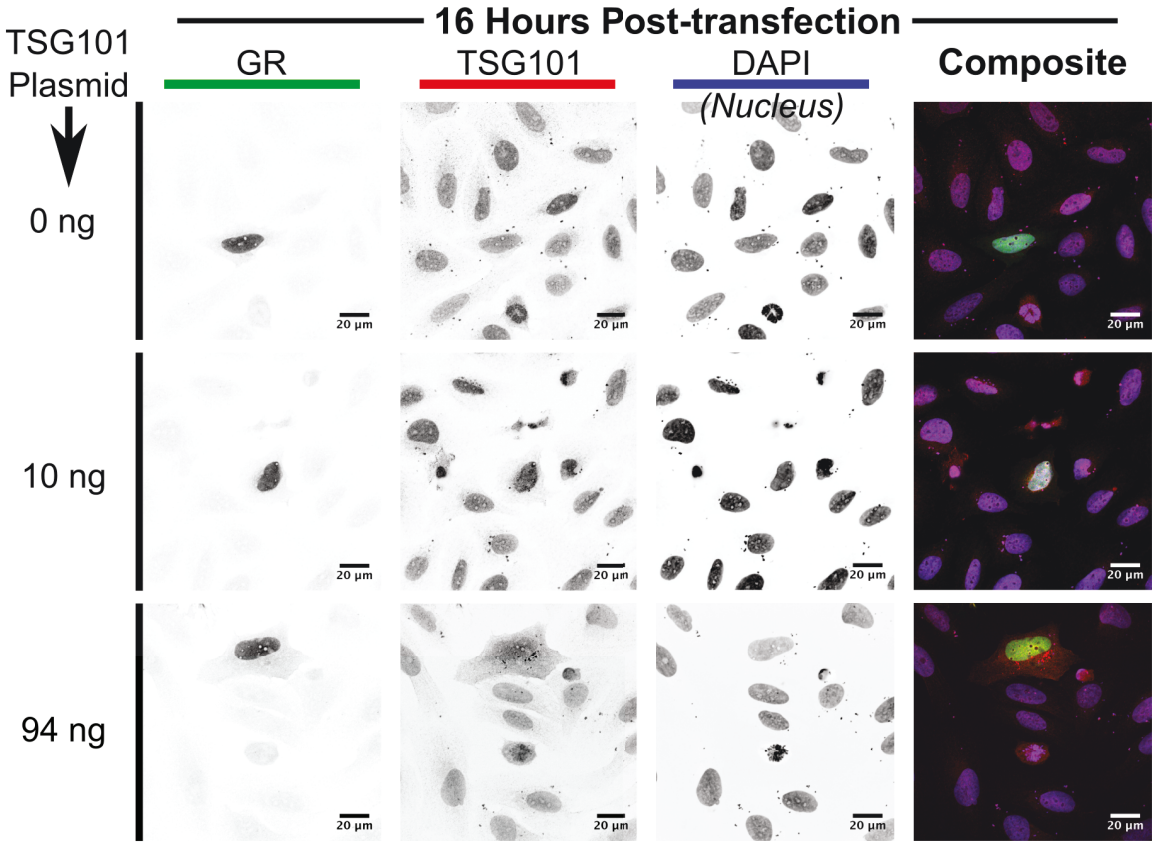


Figure 4.15 – **Overexpression of TSG101 for 16 Hours**
The left hand side describes the amount of TSG101 plasmid transfected. All scale bars are 20 µm. GR and TSG101 were detected by immunostaining. All cells were transfected with a constant 6 ng of C3 NTD-DBD plasmid.

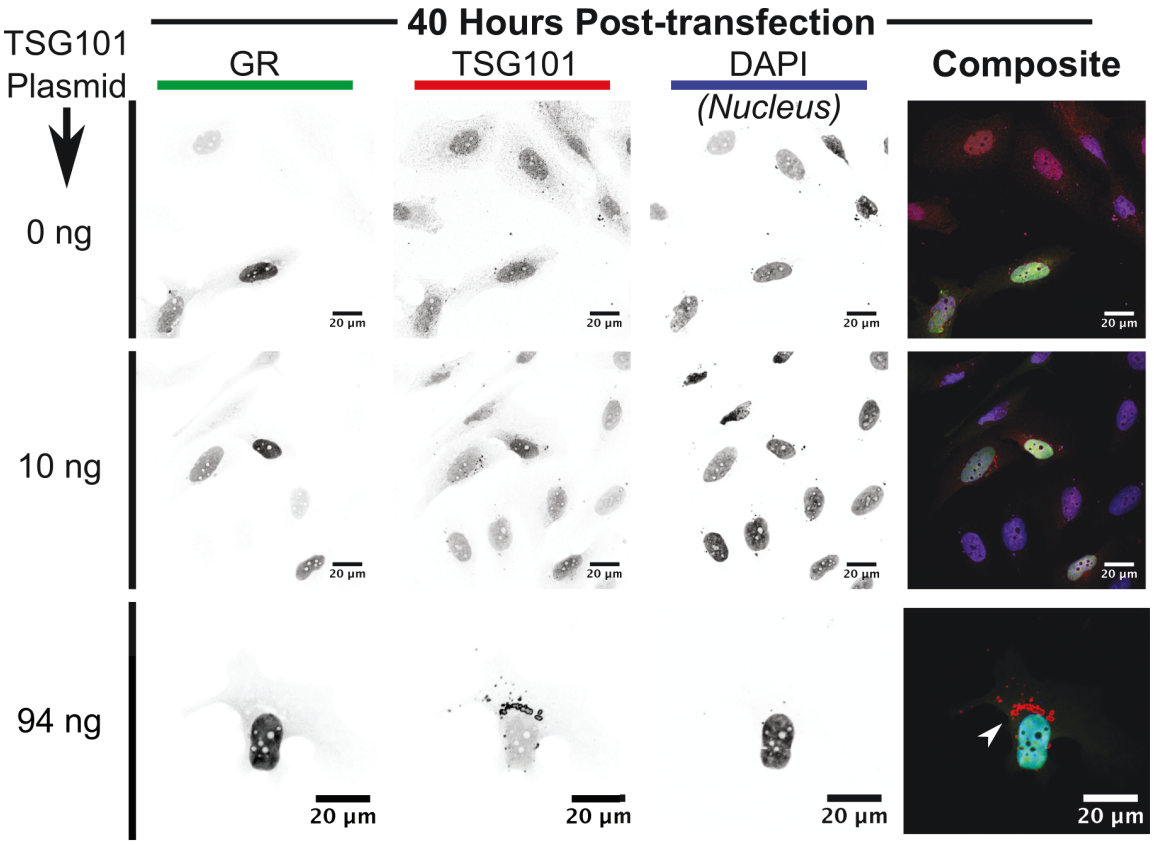


Figure 4.16 – **Overexpression of TSG101 Eventually Causes Endosomal Enlargement**
The same experiment as in Figure 4.15, except 40 hours post-transfection. All scale bars are 20 µm. Note that the bottom set of images has been zoomed in on one cell to highlight the endosomal enlargement (arrowhead).

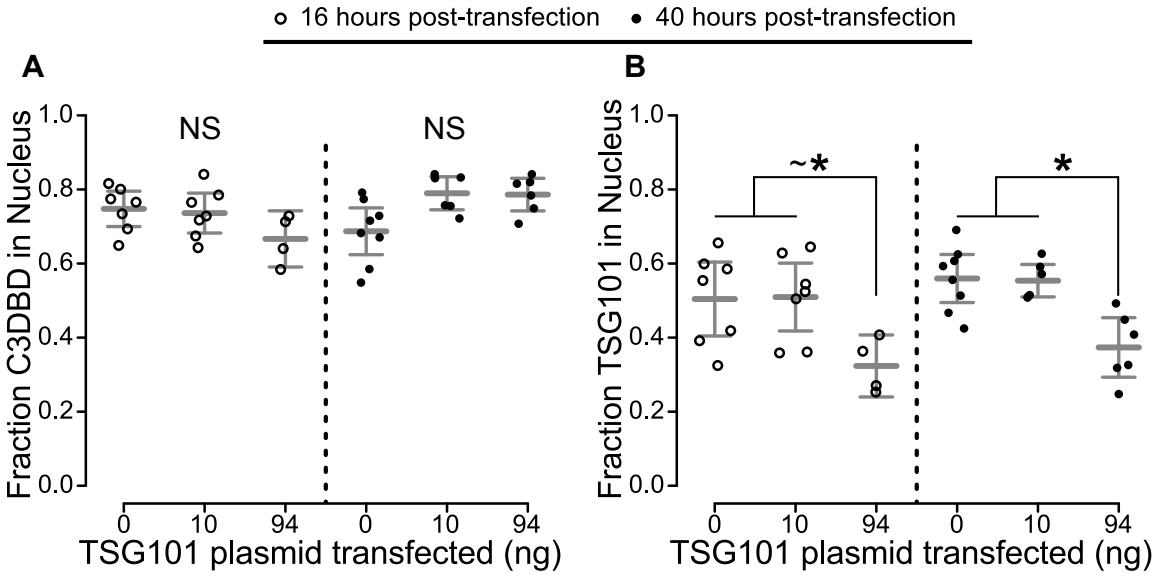


Figure 4.17 – **Localization of GR and TSG101**
Cells from the Figure 4.15 data set that coexpressed C3 NTD-DBD and TSG101 were analyzed using Fiji. Error bars are 95% confidence intervals. Statistical significance was determined by a Tukey HSD test in R: NS = not significant, ~* P ≈0.05, * P < 0.05. **A** GR has a constant and high nuclear localization. **B** TSG101 saturates its nuclear localization at high expression levels. Note that the y-axis is not in absolute but relative units.

4.4.5 TSG101 Can Increase the Transcriptional Activity of GR

Seen in Figure 4.18 transfection of increasing amounts of TSG101 plasmid increased the transcriptional activity of the active isoforms of GR (A—C3 NTD-DBD). The inactive isoforms (D1—D3 NTD-DBD) and TSG101 alone cannot activate expression of *Gaussia* luciferase. Note that the inactive isoforms of GR lack the region that binds TSG101cc and other cofactors; thus the observed effects depend on GR's innate transcriptional activity and its ability to bind TSG101cc. To test this further, I made two mutant TSG101 constructs: a coiled-coil deletion

(TSG101Δcc) or the coiled-coil on its own (TSG101cc).

The TSG101cc was sufficient to cause activation of either the A or C3 NTD-DBD isoforms of GR, while deletion of the coiled-coil ablated the effect (Figure 4.19). However, the effect of TSG101cc is much weaker than the full-length TSG101, and the activation shown in Figure 4.19 was observed in 2 out of 3 and 4 out of 5 replicates with the A and C3 isoforms of GR, respectively. I speculate that the TSG101cc likely acts to initially assemble a TSG101 transcriptional complex, but the other TSG101 domains are critical to full functionality.

Lastly, it was previously reported that TSG101 could stabilize AR and GR [63, 299], so I used Western blots to determine how TSG101 affects intra-cellular concentrations of GR in my luciferase assay. Transfection of increasing amounts of TSG101 was associated with increases in GR (C3 NTD-DBD, see Figure 4.20). Given the previously shown luciferase assays (Figure 4.18 and Figure 4.19), TSG101 stabilizes a transcriptionally active form of GR.

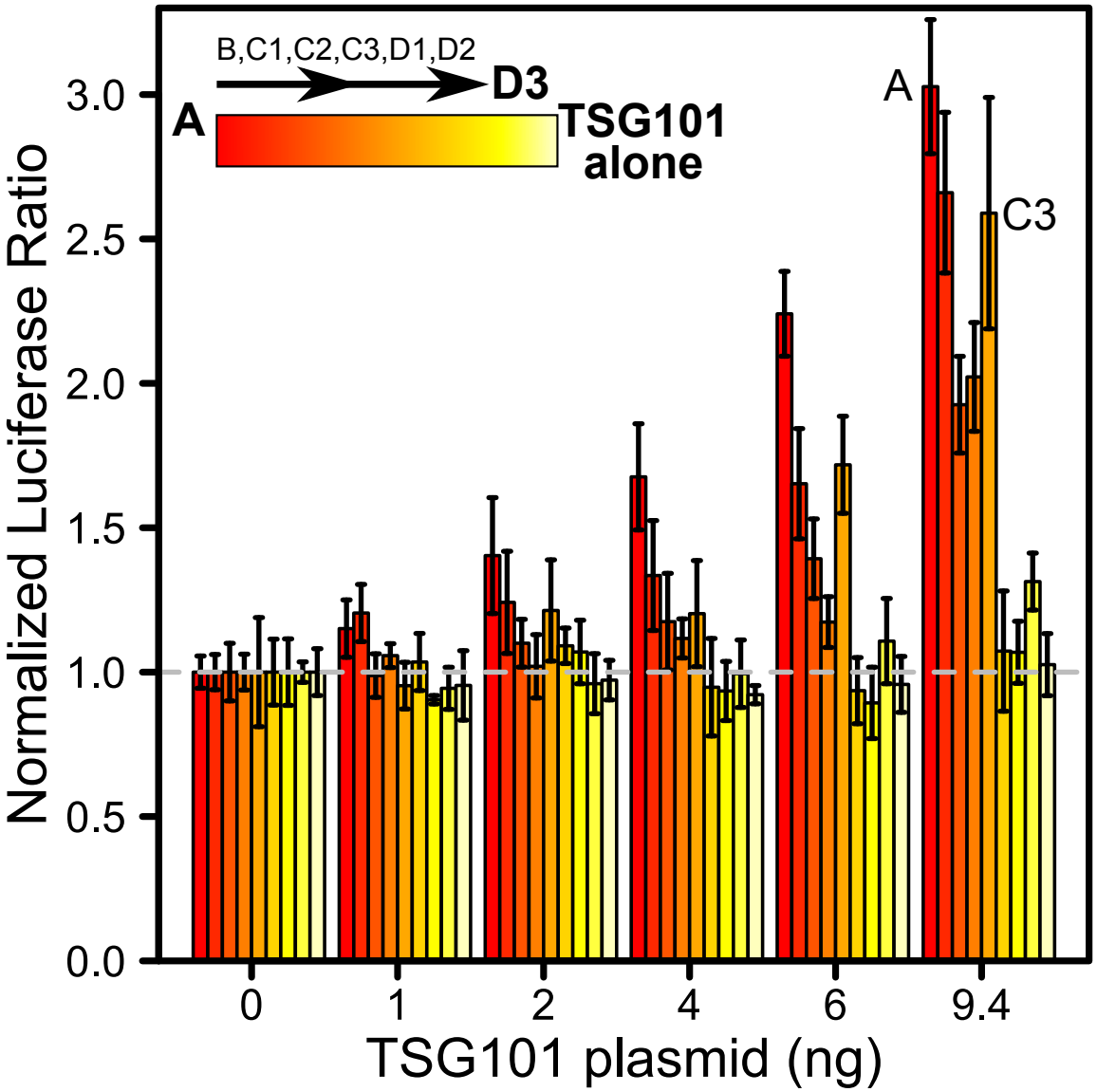


Figure 4.18 – **TSG101 Can Activate GR**
GR plasmids were transfected to 96-well plates at constant amounts as TSG101 plasmid was titrated in increasing amounts (x-axis groups). The raw luciferase ratio was normalized to signal at 0 ng (y-axis). The data are arranged with the A isoform on the far left and each consecutive isoform follows in order, as shown in the inset graphic. TSG101 was transfected by iself as a negative control and is the last bar of each group. The A and C3 bars of the 9.4 ng group are highlighted to show that the active isoforms are activated by TSG101. Error bars are mean ± 95% CI of at least four samples.

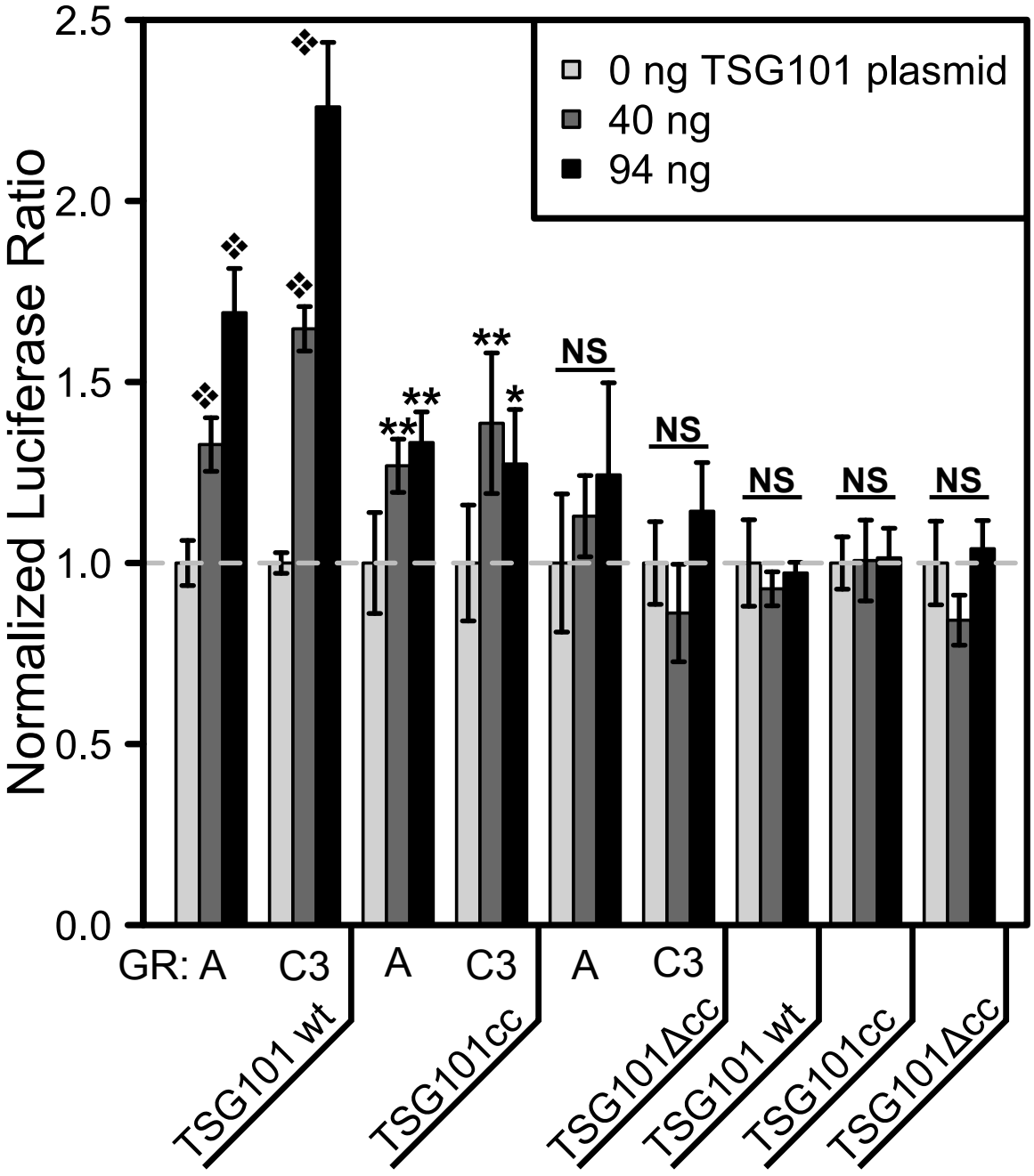


Figure 4.19 – **TSG101's Coiled-Coil is Necessary for Activation**
All data shown here were done with 6-well plates (10x amount of DNA relative to 96-well plates). For GR samples, a constant 6 ng of GR plasmid was transfected. T-tests were one-sided with the alternative hypothesis being that the 0 ng group mean < experimental mean. Benjamini-Hochberg corrections were used throughout [42]. Error bars are mean ± 95% CI of four samples. *P < 0.05, **P < 0.01, black diamond is P < 0.001

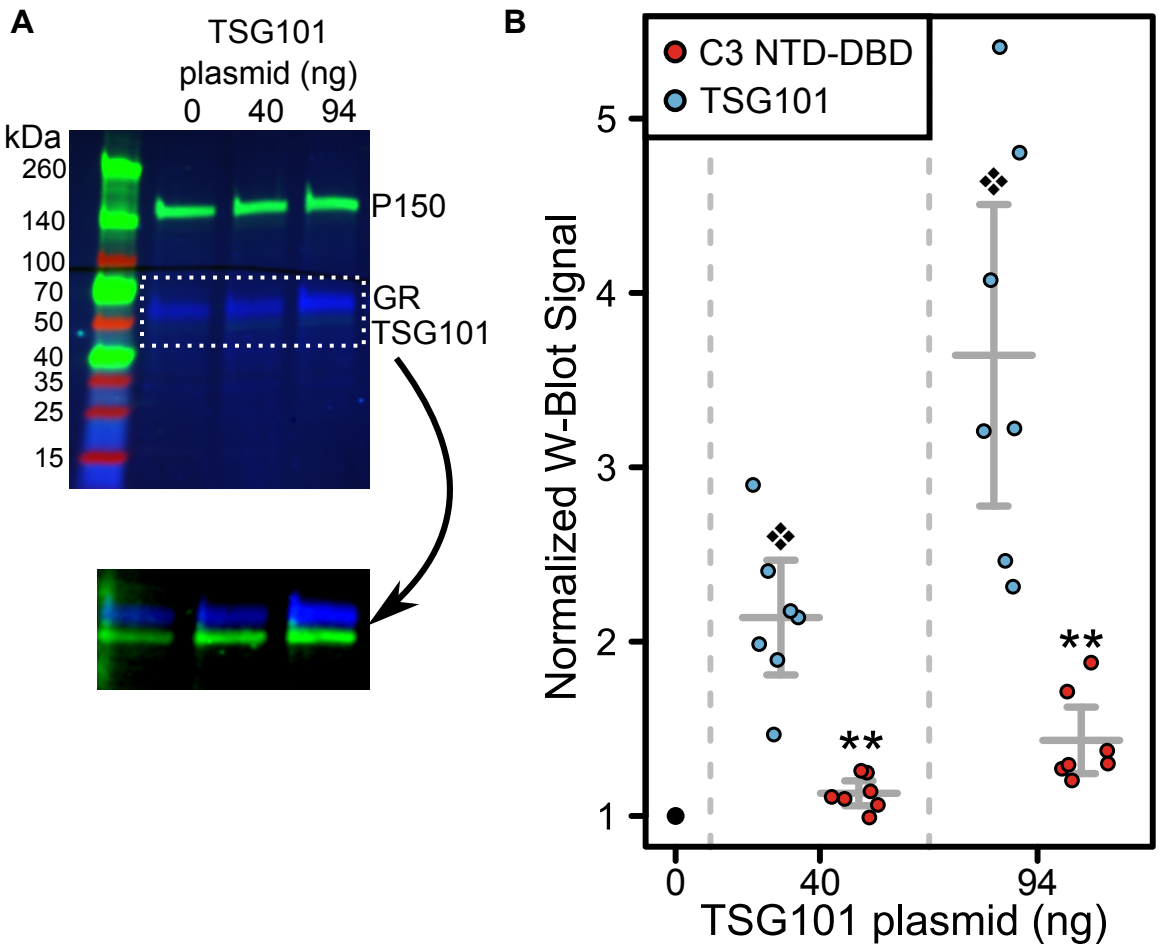


Figure 4.20 – **TSG101 Stabilizes a Transcriptionally Active Form of GR**
Cells were grown in 6-well plates and transfected with a constant 6 ng of C3-DBD while TSG101 plasmid was titrated. **A)** A representative W-blot. A contrast altered image is in the lower right. P150 was used as a loading control and the data of each lane were normalized to it and the respective 0 ng signal for **B)** Quantification of the W-blot signal. T-tests were one-sided with the alternative hypothesis being that the 0 ng group mean < experimental mean. Benjamini-Hochberg corrections were used throughout [42]. Error bars are mean ± 95% CI. **P < 0.01, black diamond is P < 0.001

4.5 Discussion

In recent years, many proteins have been discovered moonlighting away from their previously discovered roles, and this phenomenon is not unique to any particular kingdom or domain of life (reviewed: [114, 265, 429, 441]). Moonlighting has been described for metabolic enzymes (ex. PKM2 [364]), cytoskeletal proteins (ex. villin [312]), and even several ESCRT proteins, besides TSG101. In its non-endosomal roles, the ESCRT-II complex binds *bicoid* mRNA in metazoans [173] and it also forms a transcription elongation complex [192, 358]. The ESCRT-III protein, CHMP1, is involved in chromatin silencing with the polycomb group of proteins [385], and some ESCRT-I proteins have been discovered in the nucleus (VPS37C and MVB12A, [406]), though the exact function is unclear.

When mammalian TSG101 was discovered in 1996, it was proposed to be a transcription factor with its coiled-coil acting as a DNA-binding leucine zipper [244]. However, DNA-binding leucine zippers are basic [282] and TSG101's coiled-coil is acidic (pI ≈5.5, SEDNTERP by John Philo). The acidity of the TSG101cc makes it more akin to the acidic transcription-activation domains found in many transcription factors [158, 282, 335], including the GR NTD [154]. Throughout the years much research on TSG101 has focused on its ESCRT functions [23, 102, 117], but reports have trickled in of TSG101 having transcriptional effects or being present in the nucleus [293, 449, 459]. TSG101's presence in the

nucleus could be explained away by some ESCRT functions. Epstein-Barr virus hijacks the ESCRT system to exit the nucleus [74] and nuclear membrane repair has been shown to be ESCRT-III dependent [330].

Yet, even the example of Epstein-Barr virus suggests there is more to nuclear TSG101. The viral Rta protein depends on a direct interaction with TSG101 for its transcriptional activity [74]. Several other studies have linked TSG101 to either activation [63] or repression of transcription factors [153, 293, 392, 432]. Some of these studies have used yeast two-hybrids or pull-down assays to check for direct interactions [63, 74, 153, 293, 392], but a biophysical characterization has been lacking.

Here, I confirm some previous cellular observations and extend them into a quantitative realm. In U2OS cells, TSG101 was in the nucleus constitutively and it activated glucocorticoid receptor in a manner that depended on both the TSG101 binding site (TSG101cc) and the GR binding site (the C3-NTD region). Transfection of TSG101 also appeared to increase the amount of GR in the cells, similar to previous reports of TSG101 inhibiting the degradation of GR and AR [63, 299].

In vitro, protease protection patterns confirmed the TSG101cc–GR NTD interaction site that was first identified by others in a yeast two-hybrid [153], but the protected fragments were much larger than the combined mass of the fragments observed in MALDI. Cross-linking TSG101cc and C3 NTD produced

several cross-links out to a.a. 296 on GR, resolving the discrepancy in my protease protection experiments. It is worth noting that my model of TSG101cc binding to the GR NTD is consistent with the most N-terminal portion of GR (R region) being close to the DBD in the transcriptionally active state, as I previously proposed [243]. I further characterized the binding of TSG101cc and the GR NTD by CD and SMFS.

CD suggested that both TSG101cc and GR may go through conformational transitions, potentially conflating any analysis of the CD data. But in agreement with a yeast two-hybrid of GR and TSG101 [153], a large fragment of the TSG101cc (a.a. 229—339) bound the C3-NTD, and I observed a gain in α-helical structure, similar to previous reports with GR and other cofactors [201, 219]. In contrast, a fragment of the GR NTD (AF1 core) that is next to the TSG101 binding site does not seem to bind TSG101cc, as it and TSG101cc produced no change in dichroism. I next used SMFS to confirm that specifically GR was going through conformational transitions. The AF1 core was used in SMFS to control for experimental artifacts, and it had no transitions regardless the presence of TSG101cc. The C3-NTD and the AF1-binding site of TSG101 both produced transitions in the presence of TSG101cc and only one without it. This suggests that the GR NTD can fold in the presence of the TSG101cc.

To quantify the binding of TSG101 and GR, I used fluorescence anisotropy of TSG101cc and GRE, separately. The TSG101cc bound the A and C3 isoforms

of the GR NTD equally well ($K_d \cong 20$ μM). This was unexpected because the C3 isoform is thermodynamically more stable and thus predicted to fold and bind transcriptional cofactors more readily [242], which would have explained previous observations that the C3 isoform is transcriptionally more active [41, 260]. It is unclear why my hypothesis was false, but it may be related to the TSG101cc folding only a portion of the NTD. Previous Hilser lab measurements used an osmolyte to completely fold the GR NTD [242], thereas the data here suggest that TSG101cc is binding and folding only a portion of the GR NTD. Other transcriptional cofactors may bind the osmolyte-folded state and produce both the thermodynamic and transcriptional differences between the A and C3 isoforms (there are over a hundred transcriptional cofactors [255]).

GRE binding experiments demonstrated that the TSG101cc can allosterically promote DNA-binding by both the A and C3 isoforms of GR, to a similar degree. However, the exact mechanism differs between the two isoforms. Experiments with the GRE half-site revealed that TSG101cc promotes a DNA-binding state of the C3 isoform's DBD, but promotes the dimerization energy of the A isoform. This suggests that TSG101 could affect the DBD conformational ensemble differently depending on the GR isoform. I speculate that such conformational heterogeneity could cause the TSG101xGR interaction to be targeted to specific DNA elements in an isoform-dependent manner. It has already been shown that the GR translational isoforms affect unique but overlapping sets of genes

[260], and binding of different DNA sequences is associated with different DBD conformations [278] (conformational heterogeneity is important for the specificity of other DNA-binding proteins too [4, 191, 379]). Furthermore, GRE sequence has already been shown to affect binding of another transcriptional cofactor of GR, BATF3 [44].

If the TSG101 interaction with steroid hormone receptors depends on the promoter that the receptor binds, then it would explain some of the effects seen in the literature. Numerous authors have reported TSG101 both activating and repressing various transcription factors (see Table 4.1 and associated references). Different labs have used cognate response elements or various viral promoters to measure the effect of TSG101 on steroid hormone receptors [63, 74, 293, 432]. Besides using different reporter assays, the same labs also used a wide variety of cell lines, which may have expressed varying amounts of cofactors important for TSG101's interaction with steroid receptors (Table 4.1). In particular, the transcription factor Daxx and the cofactor AATF appear to modulate the effect of TSG101 on GR and AR, respectively [63, 293]. At the moment it is not clear what has produced conflicting results (activation versus repression) in the literature, and I believe that this matter warrants further study.

Figure 4.21 summarizes my results and hypothetical model of TSG101cc binding to GR based on the above observations and literature data. In this study, I have defined TSG101 as a transcriptional cofactor that stabilizes an

active, DNA-binding form of GR *in vivo*. Open questions include which genes are naturally controlled by this interaction and how this ties in, if at all, with ESCRT-II's transcriptional activity. My quantification of the allosteric energies present in GR enables future studies to build organizing principles out of this system's thermodynamics.

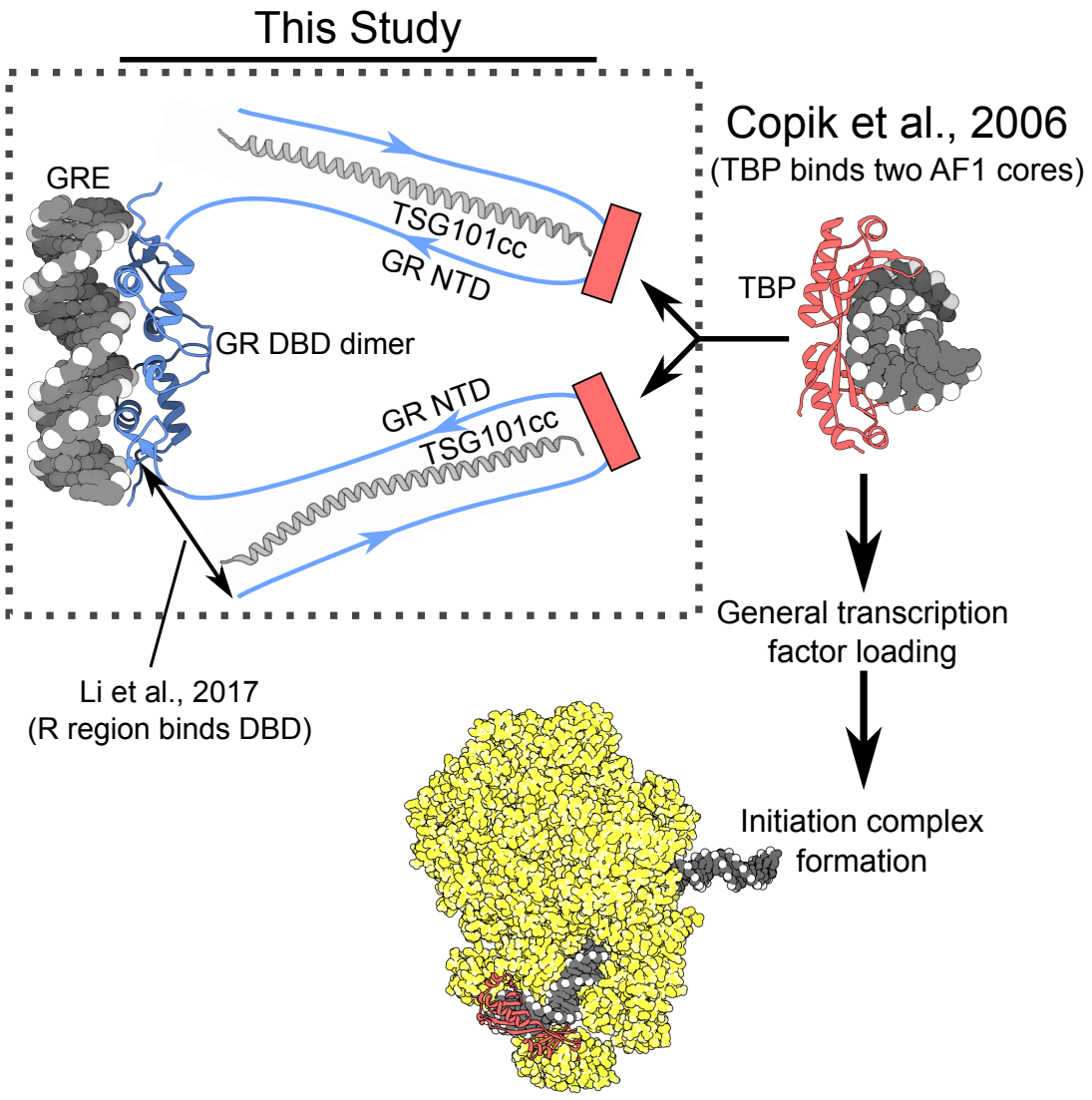


Figure 4.21 – TSG101 May Directly Activate GR
The results of this study are schematized in the boxed in area. GR (blue) bound to its GRE (grey) is shown as before. Each GR NTD forms a pseudo-trimer with one TSG101cc (grey coils, pdb:3iv1). The individual GR NTD's do not interact (our data only show heterodimers of GR and TSG101cc). Also note that the Hilser lab previously showed that the R region of GR (A NTD) binds the DBD and this schematic model allows for that association. TSG101's interaction with GR does not depend on the TBP binding site (pink), possibly leaving it open for interaction with TBP. On the **right**, TBP is shown in pink. Others demonstrated that TBP binds two AF1 core elements of GR [75]. The series of arrows represent subsequent recruitment events, eventually producing an active transcriptional complex (yellow). Structures: [141, 263, 302]

4.6 Appendix

4.6.1 Tables of MS Ions Observed

Table 4.4 – Ions Detected in MS of Protease Protection
The left two columns are for full-length C3-NTD that was extracted from the gel. The right two columns are for the fragments of GR protected by TSG101cc. Italics indicates a fragment that was excluded from the dataset shown in the main text because it could ambiguously be from TSG101cc. Note that the fragment is in the GR alone dataset, though. All amino acid numbers are relative to the holo-protein sequence.

C3NTD alone:		+TSG101cc	
ion ± error	a.a.	ion ± error	a.a.
1018.46 + 0.03	*207-214*	*1018.46 + 0.07*	*207-214*
1032.5 - 0.09	172-184	1032.5 - 0.06	172-182
1271.7 + 0.05	121-131	1495.7 + 0.01	159-171
1495.7 + 0.04	159-171	1807.9 + 0.01	141-158
1807.9 + 0.01	141-158	2211.1 - 0.03	121-140
1821.9 -0	278-293	2649.3 + 0.01	184-207
2049.0 + 0.01	367-386		
2224.1 - 0.01	257-277		
2479.18 - 0.03	97-119		
2649.2 - 0.02	183-206		
3310.6 + 0.02	387-419		

Table 4.5 – **Ions Detected in MS of Unreacted C3NTD and TSG101cc**
C3NTD data are on top and TSG101cc is below. An ambiguous ion is italicized to show that it exists in both GR and TSG101 MS. All amino acid numbers are relative to the tag-cleaved construct.

C3NTD unreacted			
Trypsin		Tryp+Chymotryp	
ion ± error Da	a.a.	ion ± error Da	a.a.
1271.695 - 0.37	25-35	787.36 + 0.35	291-297
1495.71 + 0.78	63-75	958.48 - 0.07	36-44
		1018.46 - 0.06	*111-118*
		1271.695 - 0.02	25-35
		1495.71 - 0.01	63-75
		1807.01 - 0.3	45-62
		2088.98 - 0.39	76-95
		2224.22 - 0.43	161-181
		2246.11 - 0.47	36-58

TSG101cc unreacted			
Trypsin		Tryp+Chymotryp	
ion ± error Da	a.a.	ion ± error Da	a.a.
597.34 + 0.13	39-43	597.35 -0.17	39-43
729.45 - 0.04	61-66	729.45 - 0.01	61-66
862.45 - 0.07	33-39	734.36 + 0.01	33-38
877.44 + 0.03	44-50	877.45 - 0.07	44-50
928.4 +0	16-22	938.41 - 0.12	16-22
957.52 +0.02	23-31	*1018.55 - 0.09*	*32-39*
1018.55 + 0.02	*32-39*	1063.56 - 0.17	1-11
1063.56 - 0.01	1-11	1145.57 - 0.04	51-60
1113.64 + 0.02	23-32	1248.63 - 0.09	68-78
1248.63 - 0.07	68-78		
1332.75 - 0.07	1-13		
1376.73 - 0.08	67-78		
1856.0 + 0.04	51-66		

Table 4.6 – **Ions Consistent with GR or TSG101 in MS of Reacted C3NTD and TSG101cc #1**
C3NTD data are on top and TSG101cc are below. An ambiguous ion is italicized to show that it exists in both GR and TSG101 MS. All amino acid numbers are relative to the tag-cleaved construct.

C3NTD DST reacted			
Trypsin		Tryp+Chymotryp	
ion ± error Da	a.a.	ion ± error Da	a.a.
520.28 - 0.59	59-62	552.30 + 0.03	201-205
926.46 + 0.12	324-330	787.36 + 0.08	291-297
958.48 + 0.17	36-44	926.46 -0.02	324-330
1018.46 + 0.22	*111-118*	958.48 - 0	36-44
1271.7 + 0.34	25-35	*1018.46 - 0.07*	*111-118*
1306.65 + 0.38	45-58	1271.7 - 0.13	25-35
1495.71 +0.46	63-75	1495.71 - 0.12	63-75
1807.91 + 0.57	45-62	1592.78 - 0.04	96-110
2224.22 + 0.6	161-181	1618.75 - 0.13	104-118
2246.12 + 0.55	36-58	1711.78 - 0.03	220-236
2649.26 + 0.46	87-110	1807.91 + 0.02	45-62
		2224.22 + 0.69	161-181

TSG101cc DST reacted monomer			
Trypsin		Tryp+Chymotryp	
ion ± error Da	a.a.	ion ± error Da	a.a.
957.54 - 0.04	23-31	597.35 - 0.11	39-43
938.41 - 0.085	16-22	679.27 + 0.14	18-22
877.45 + 0.03	44-50	729.45 + 0.2	61-66
862.45 - 0.02	33-39	734.36 + 0.23	33-38
729.45 - 0.01	61-66	862.45 + 0.34	33-39
597.35 + 0	39-43	877.45 +0.37	44-50
1856.0 +0.77	51-66	938.41 + 0.45	16-22
1376.73 +0	67-78	*1018.55 + 0.55*	*32-39*
1332.75 - 0.07	1-13	1063.56 + 0.63	1-11
1248.63 - 0.06	68-78	1113.64 +0.53	23-32
1113.64 + 0.05	23-32	1248.63 + 0.57	68-78
1063.56 - 0.07	1-11	1327.68 + 0.61	40-50
1018.55 - 0	*32-39*	1376.73 + 0.61	67-78
		1856.0 + 0.72	51-66

Table 4.7 – Ions Consistent with GR or TSG101 in MS of Reacted C3NTD and TSG101cc #2
TSG101cc dimer data are at the top and trimer is below. An ambiguous ion is italicized. Of the higher order oligomers, only the trimer + trypsin experiment yielded good mass spectra. All amino acid numbers are relative to the tag-cleaved construct.

TSG101cc DST reacted dimer			
Trypsin		Tryp+Chymotryp	
ion ± error Da	a.a.	ion ± error Da	a.a.
957.54 - 0.04	23-31	862.45 + 0.34	33-39
938.41 + 0.05	16-22	877.45 +0.37	40-50
877.45 + 0.05	44-50	938.41 + 0.45	16-22
862.45 + 0.01	33-39	*1018.55 + 0.55*	*32-39*
729.45 -0	61-66	1113.64 + 0.53	23-32
597.35 - 0.02	39-43	1327.68 + 0.61	40-50
1856.0 +0.04	51-66	1376.73 + 0.614	67-78
1376.73 +0.02	67-78	1856.0 + 0.72	51-66
1332.75 - 0.07	1-13		
1248.63 - 0.06	68-78		
1113.64 - 0.07	23-32		
1063.56 - 0.08	1-11		
1018.55 + 0	*32-39*		

TSG101cc DST reacted trimer	
Trypsin	
ion ± error Da	a.a.
729.45 + 0.1	61-66
877.45 + 0.5	44-50
938.41 + 0.56	16-22
957.54 + 0.41	23-31
1114.64 + 0.61	23-32
1856.0 +0.58	51-66

Table 4.8 – TSG101cc X-links to Itself
Amino acid numbers here refer to the tag-cleaved construct. Some of the X-links have ambiguous lysine linkages and are italicized. For reconcilable lysines, the closest Euclidean distances have been calculated directly from pdb:3iv1. DST has a maximum linker length of ∼6 Å [130]; thus, at least some of the x-links could only occur with conformational heterogeneity in the TSG101cc tetramer, as I have published before [438].

TSG101cc DST reacted X-links					
Trypsin			Tryp+Chymotryp		
ion ± error Da	a.a.	Dist. Å	ion ± error Da	a.a.	Dist. Å
1249.7 - 0.52	*16-17x61-67*		*1249.7 - 0.51*	*16-17x61-67*	
1327.62 + 0.67	*16-22x67-68*		*1327.62 + 0.7*	*16-22x67-68*	
1894.92 - 0.06	32-38x32-38	3.2	1728.82 - 0.25	40-43x51-60	33.3
2778.3 + 0.18	32-38x18-32	10.8	*1765.93 + 0.39*	*67-68x67-78*	
			2510.31 - 0.03	1-11x1-13	25.9

Table 4.9 – **Ions of C3NTD and TSG101cc Observed After X-linking**
These data are from the TSG101cc-C3NTD cross-link band. On the left are ions consistent with GR or TSG101cc, unreacted. Cross-linked amino acids are listed relative to the holo-protein sequence. All other amino acids are relative to the tag-cleaved construct. One TSG101cc mono-link was observed: 601.25 + 0.07 Da and a.a. 267-270 of holo-protein sequence.

C3NTD+TSG101cc DST X-link Ions		
Trypsin		
ion ± error Da	TSG a.a.	GR a.a.
1475.75 + 0.04	250-258	294-296

Trypsin+Chymotrypsin		
ion ± error Da	TSG a.a.	GR a.a.
1329.65 - 0.21	243-244	294-301
1448.74 - 0.11	250-258	278-280
1615.76 - 0.01	267-270	172-182
1744.79 + 0.04	260-266	278-283

4.6.2 TSG101 Evolutionary Alignment

Figure 4.22 – Amino Acid Alignment of Human TSG101 and Ten Homologues (next three pages)
Alignment was done using MUSCLE [101]. The species are presented as three letter acronyms on the left of each stretch of sequence: T.th = *Tetrahymena thermophila*, T.br = *Trypanosoma brucei*, S.ce = *Sacharomyces cerevisiae*, V.vi = *Vitis vinifera* (wine grape), C.el = *Caenorhabditis elegans*, D.me = *Drosophila melanogaster*, X.tr = *Xenopus tropicalis*, D.re = *Danio rerio*, G.ga = *Gallus gallus*, H.sa = *Homo sapiens*, M.mu = *Mus musculus*. At the bottom of each text block are marks indicating conservation of sequence, where "." is weakly similar, ":" is very similar, and "*" is perfect identity. Note that there are many sections where perfect conservation with human TSG101 is broken by only a few of the sequences. Perfect conservation with the human sequence is highlighted in yellow within the coiled-coil region of the alignment.

```
T.th  -------------MNLQQICSQ-------C--QYKNFLQVIQAGQELIN------QGFIA
T.br  -----MNDRELRAYTKLDVLKR----------HYSA--LAAEALHDYINDFMPYTENFKN
S.ce  --MSANGKISVPEAVVNWLFKV-------IQPIYNDGRTTFHDSLALLDNFHSLRPRTRV
V.vi  MASSSSSPPQSSHQFLTSVLSRRGPSALPYAEDVKW--MIRQHFLSLADAFPALHPQTAA
C.el  ----------MSAHQVQQCLQR-------AGGKYAD--SAKKDIIGALSQFKDLSPGTDT
D.me  -------MPAVEETQITKYLSK-------Y--KYVA--ATKKDVVDVVTSFRSLTYDLQR
X.tr  --------MALSEAQLKKLLGK-------Y--KYRD--LSVREILNVTSLYRDLKPIMDS
D.re  -------MAVVNEGALKKMLKQ-------Y--KYRD--LTVREITNVISQYKDLKPVMDA
G.ga  --------MAFSESQLKKMLAK-------Y--KYRD--LTIQETNSVISQYKDLKPVMDS
H.sa  --------MAVSESQLKKMVSK-------Y--KYRD--LTVRETVNVITLYKDLKPVLDS
M.mu  --------MAVSESQLKKMMSK-------Y--KYRD--LTVRQTVNVIAMYKDLKPVLDS

T.th   SMYQQKIGNQLDSVLIMSGFV----------LVNYQQQP-LQINTEVIFSKEYPISPPAF
T.br   ISCHFCNNDKQL-LLTLDCPIYVMPPGVNAAAAPRQGLV-HPLRVEVVFLANFPNSAP-T
S.ce   --FTHSDGTPQL-LLSIYGTI----------STGEDGSSPHSIPVIMWVPSMYPVKPP-F
V.vi   --YTHNDGRTVN-LLRVEGNI----------PMVYLQVT-YYIPIVVWLMEAYPRQPP-C
C.el   --FMFPDGKRRT-AFRLKGTI----------PVYYKGAC-YNIPVTVYLWDTHPYYAP-I
D.me   --FVFNDGSSKE-LFTIQGTI----------PVVYKNNT-YYIPICIWLMDTHPQNAP-M
X.tr   --YVFNDGSSRE-LLSLSGTI----------PVPYKGNT-YNIPICLWLLDTYPFNPP-I
D.re   --YVFNDGSSRD-LMSLTGTV----------PVSYRGNV-YNIPVCLWLLDTYPYNPP-I
G.ga   --YVFNDGSSRE-LMSLSGTI----------PVPYRGNV-YNIPICLWLLDTYPFNPP-I
H.sa   --YVFNDGSSRE-LMNLTGTI----------PVPYRGNT-YNIPICLWLLDTYPYNPP-I
M.mu   --YVFNDGSSRE-LVNLTGTI----------PVRYRGNI-YNIPICLWLLDTYPYNPP-I
              .      . :   :                    :   : .   .*  .*

T.th   YFVNP-------DPNKFQVNSKY-VINKRPNSESYYVILDNLN-WNT---HKDLKKTMNL
T.br   VKLLPNTFRCGNLTEWRVRPASS-IMDED-----GAIHLEKLALLEQAVVPYSLLEILVA
S.ce   ISINLENFDMNTISSSLPIQE---YIDSN-----GWIALPILHCWDPA--AMNLIMVVQE
V.vi   VYVNP--------TRQMVIKRPHPHVNPS-----GLVSLPYLQNWIYP--TSNLADLARN
C.el   CYVNP--------TSTMVIKESE-HVNKE-----GKVFLPYLNEWRFP--GYDLSGLLQV
D.me   CFVKP--------TPTMQIKVSM-YVDHN-----GKVYLPYLHDWQPH--SSDLLSLIQV
X.tr   CFVKP--------TSTMTIKTGK-HVDAN-----GKIYLPYLHEWKHP--PSDLLGLIQI
D.re   CFVKP--------TSAMMIKTGK-HIDAN-----GKIYLPYLHEWKHP--QSDLYGLIQV
G.ga   CFVKP--------TSSMTIKTGK-HVDAN-----GKIYLPYLHEWKYP--QSDLLELIQV
H.sa   CFVKP--------TSSMTIKTGK-HVDAN-----GKIYLPYLHEWKHP--QSDLLGLIQV
M.mu   CFVKP--------TSSMTIKTGK-HVDAN-----GKIYLPYLHDWKHP--RSELLELIQI
         :          .          ::         : *  *          .*
```

Figure 4.22

```
T.th   YIQILQKDFPLYSK----------------------------QFQNQNNQKNYEIN----
T.br   LTEQFEAEFP--------------------------------LTPSEREPIGQNPP----
S.ce   LMSLLHEPPQ-------DQAPSLP------------------PKPNTQLQQEQNTP----
V.vi   LSLVFGRDPPLYSQQ---------------------------PRPNLNPNS---IP----
C.el   MAMVFQEKCPVFARS--AANSATNASATNPSAGSSASSTP-TPYPSSQPT----MPTPYP
D.me   MIVTFGDHPPVYSKPKEQIAAPYPTNSYMPQPGAPGGSNSFLPYPTAGGAGGSNFP-PYP
X.tr   LIVVFGEEPPVFSRS--TAPAPYP------------------MYPATGPPNSSYMP-GV-
D.re   MIVVFGEEPPVFSRP--TTQPPYQ------------------AFQATGPPNQSYMP-GMP
G.ga   MIVVFGEEPPVFSRP--TVSASYP------------------PYQATGPPTTSYVP-GMP
H.sa   MIVVFGDEPPVFSRP---ISASYP------------------PYQATGPPNTSYMP-GMP
M.mu   MIVIFGEEPPVFSRP--TVSASYP------------------PYTATGPPNTSYMP-GMP
           :

T.th   ----------------NQNQIG-----------------GDIEQIIYN------------
T.br   ---IATATPP---VSSIVNSVVNSTFC----PTDGSAP-----PNVMPVPEMPTSDPSVV
S.ce   ------PLPP-----KPKSPHLKPPLP----PPPPPQP-ASNALDLMD------MDNTDI
V.vi   --------------------------------PRA------------FR-------PSSPH
C.el   TGSGAAPYPPSSTPYPSAGAMGYNPYMNV--PQSTPYPMGASGASPYPSASSNPAPPPPR
D.me   TGSNVGPYPP--TPAGPAGGSGYPAYPNFIQPTAGGYP-PAAG---YN----------PS
X.tr   ----IPPYPP---AAHPANPSGYGGFPGY--PPAGQYP-QTSGPQIFPQPPAAQPPVTSS
D.re   -A--VSPY------GPNQNPCGYPGYP-Y--PGGNAYP-ATGGPSHYT----SQTPVTTV
G.ga   -G--ISPYPT----GSTANPS-FPSYP-Y--PAGVPFP-PTTNVPYYP----SQPPVTTV
H.sa   GG--ISPYPS----GYPPNPSGYPGCP-Y--PPGGPYP-ATTSSQ-YP----SQPPVTTV
M.mu   SG--ISAYPS----GYPPNPSGYPGCP-Y--PPAGPYP-ATTSSQ-YP----SQPPVTTV
```

Figure 4.22

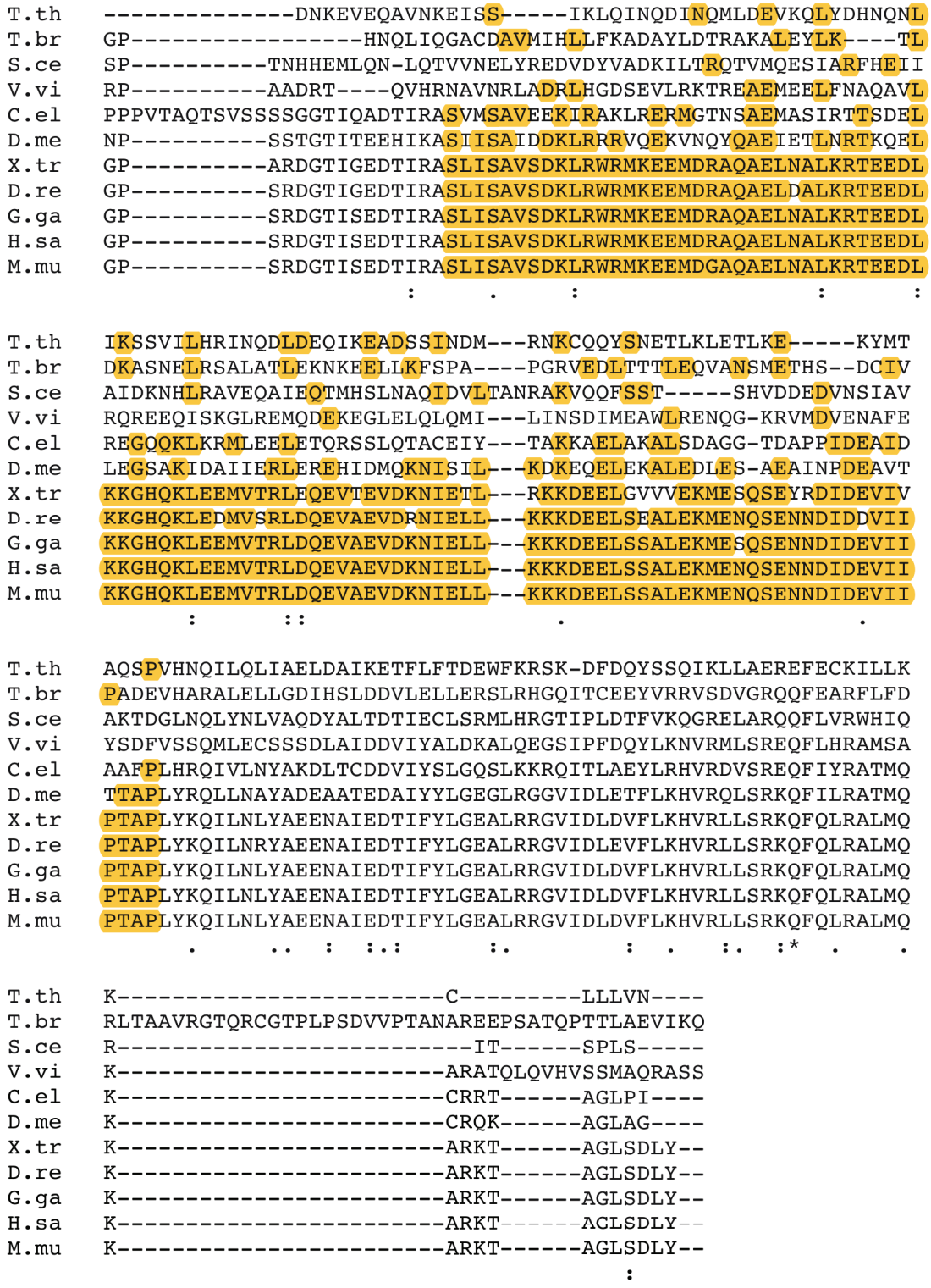


Figure 4.22

4.6.3 Contour Length Analysis for C3-NTD+TSG101cc

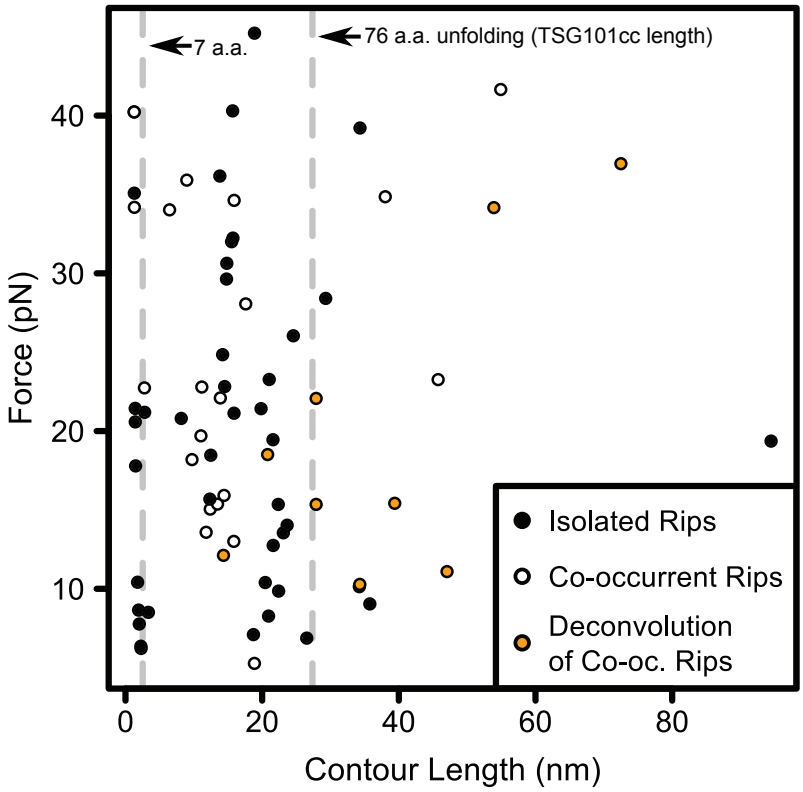


Figure 4.23 – Compilation of C3-NTD Force Ramp Data
The Worm-like Chain model was used to extract the contour lengths shown here [64]. C3-NTD was in the presence of at least 1 μM TSG101cc for all data shown here. Isolated rips occured alone in one pulling cycle, while co-occurent rips are multiple rips from one cycle. Co-occurent rip contour lengths can be added and the force can be back calculated, which is shown as the orange "Deconvolution" points. The largest contour length shown is ∼262 a.a., which is less than the C3-NTD length (323 a.a.) and longer than the TSG101cc construct used (78 a.a.).

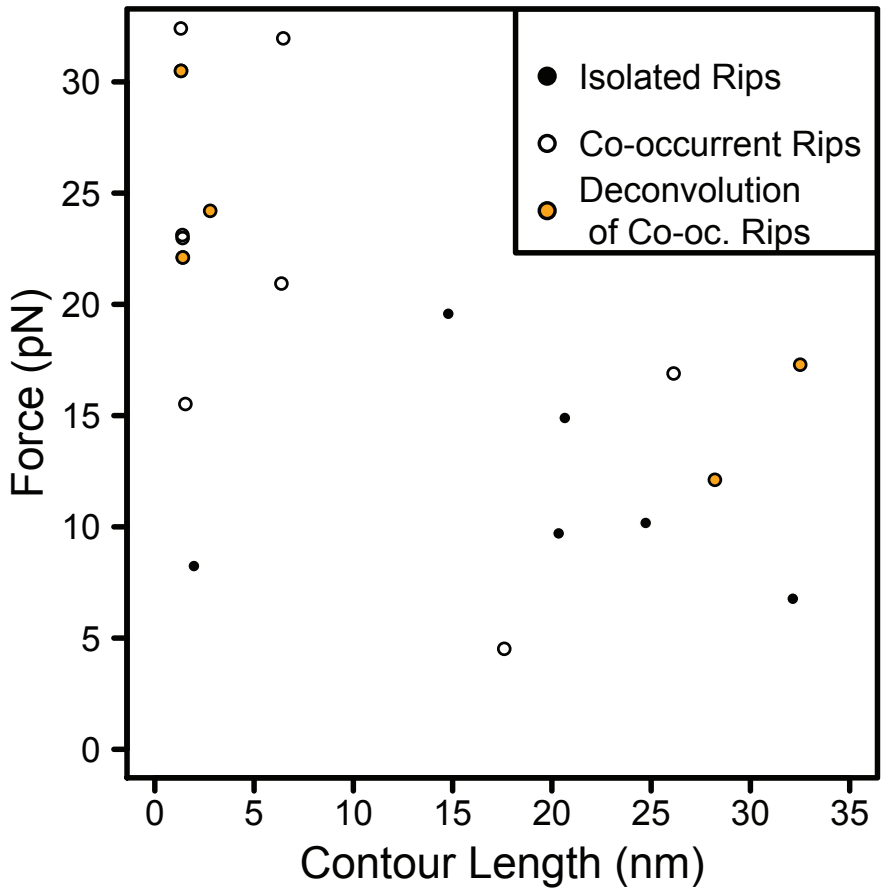


Figure 4.24 – **Compilation of AF1$^{107-237}$ Force Ramp Data**
Data are plotted as above (Figure 4.23). AF1$^{107-237}$ was in the presence of 10 μM TSG101cc for all data shown here. The largest contour length shown is ~96 a.a., which is less than the AF1$^{107-237}$ length (131 a.a.).

4.6.4 Raw Data of 6FAM-DNA Fluorescence Anisotropy

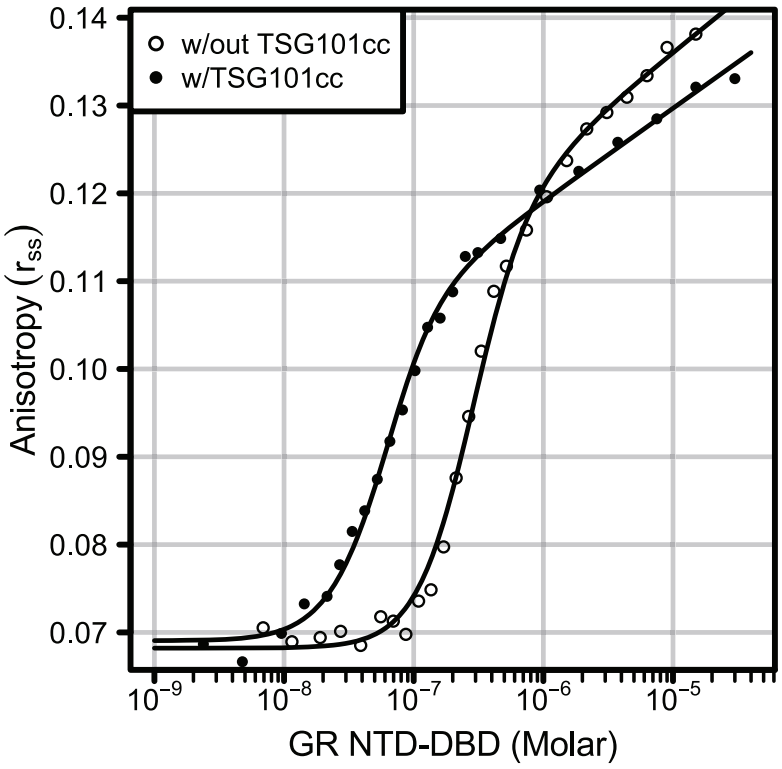


Figure 4.25 – **TSG101cc Does Not Directly Bind DNA**
The C3 NTD-DBD construct of GR was titrated out of a solution of full-length GRE, labeled with 6-FAM (observed anisotropy). For the sample with TSG101cc, 100 μM TSG101cc was maintained throughout the entire titration. Note that the unbound DNA baselines around 10^{-8} Molar are the same with or without TSG101cc.

4.6.5 Nucleotide and Amino Acid Sequences of Plasmids Cloned in This Work

4.6.5.1 Bacterial Expression Construct (pJ411 plasmid)

TSG101cc from a.a. 209 to 339, open reading frame (ORF):

ATGCACCACCACCACCATCATCACCATCACGAAAACTTATACTTCCAGGGTAGTTCTCTCATCT
CTGCGGTCAGTGACAAACTGAGATGGCGGATGAAGGAGGAAATGGATCGTGCCCAGGCAGAGCTC
AATGCCTTGAAACGAACAGAAGAAGACCTGAAAAAGGGTCACCAGAAACTGGAAGAGATGGTTACC
CGTTTAGATCAAGAAGTAGCCGAGGTTGATAAAAACATAGAACTTTTGAAAAAGAAGGATGAAGAAC
TCAGTTCTGCTCTGGAAAAAATGGAAAACCAGAGTGAAAATAACGACATTGATGAAGTGATCATCCC
GACGGCCCCGCTGTATAAACAAATCCTCAACCTGTATGCAGAAGAAAACGCAATCGAATGA

Translation of above (TEV cleavage will leave GS from tag):

MHHHHHHHHHENLYFQGSSLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLE
EMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIE

AF1τ (GR a.a. 187—244) for CD, ORF:

ATGCACCACCACCACCATCATCACCATCACGAAAACTTATACTTCCAGGGTGACCAGAGCACG
TTCGATATTCTGCAGGACCTGGAATTCAGCAGCGGCTCGCCGGGTAAAGAAACTAACGAAAGCCCG
TGGCGTTCGGATCTGTTGATCGACGAAAACTGTCTGCTGAGCCCGCTGGCCGGTGAGGACGATAG
CTTTTTGCTGGAGGGTAATAGCAACGAGGACTAA

Translation of above:

MHHHHHHHHHENLYFQGDQSTFDILQDLEFSSGSPGKETNESPWRSDLLIDENCLLSPLAGE
DDSFLLEGNSNED

AF1τ for single molecule force spectroscopy, ORF:

ATGCACCACCACCACCATCATCACCATCACGAAAACTTATACTTCCAGGGTCTGAATGATATCT

TTGAGGCTCAAAAGATTGAGTGGCATGAGGACCAGAGCACGTTCGATATTCTGCAGGACCTGGAAT
TCAGCAGCGGCTCGCCGGGTAAAGAAACTAACGAAAGCCCGTGGCGTTCGGATCTGTTGATCGAC
GAAAACTGTCTGCTGAGCCCGCTGGCCGGTGAGGACGATAGCTTTTTGCTGGAGGGTAATAGCAAC
GAGGACGATAGCCTGGAGTTCATTGCTTCTAAATTGGCATAA

Translation of above AF1τ single molecule construct:

MHHHHHHHHHHENLYFQGLNDIFEAQKIEWHEDQSTFDILQDLEFSSGSPGKETNESPWRSDL
LIDENCLLSPLAGEDDSFLLEGNSNEDDSLEFIASKLA

GR AF1 used by Hittelman (a.a. 107—237) for single molecule force spectroscopy, ORF:

ATGCACCACCACCACCATCATCACCATCACGAAAACTTATACTTCCAGGGTCTGAATGATATCT
TTGAGGCTCAAAAGATTGAGTGGCATGAGCAGGGCCAGATTTCACTGAGCAGCGGCGAAACCGACT
TGAAACTCCTGGAAGAGAGCATTGCGAACCTGAATCGTAGCACTTCCGTTCCGGAGAACCCGAAGA
GCAGCGCGTCGACTGCCGTTAGCGCAGCGCCGACGGAGAAAGAATTTCCGAAAACTCACTCCGAC
GTAAGCAGCGAGCAGCAACACCTGAAGGGCCAGACCGGCACTAACGGTGGTAACGTGAAACTGTA
CACGACCGACCAGAGCACGTTCGATATTCTGCAGGACCTGGAATTCAGCAGCGGCTCGCCGGGTA
AAGAAACTAACGAAAGCCCGTGGCGTTCGGATCTGTTGATCGACGAAAACTGTCTGCTGAGCCCGC
TGGCCGGTGAGGACGATAGCTTTTTGCTGGATAGCCTGGAGTTCATTGCTTCTAAATTGGCATAA

Translation of above Hittelman AF1 single molecule construct:

MHHHHHHHHHHENLYFQGLNDIFEAQKIEWHEQGQISLSSGETDLKLLEESIANLNRSTSVPEN
PKSSASTAVSAAPTEKEFPKTHSDVSSEQQHLKGQTGTNGGNVKLYTTDQSTFDILQDLEFSSGSPG
KETNESPWRSDLLIDENCLLSPLAGEDDSFLLDSLEFIASKLA

C3-NTD for single molecule force spectroscopy, ORF:

ATGCACCACCACCACCATCATGAAAACTTATACTTCCAGGGTCTGAATGATATCTTTGAGGCTC
AAAAGATTGAGTGGCATGAGATGGGTAATGATCTGGGCTTCCCTCAGCAGGGCCAGATTTCACTGA
GCAGCGGCGAAACCGACTTGAAACTCCTGGAAGAGAGCATTGCGAACCTGAATCGTAGCACTTCC
GTTCCGGAGAACCCGAAGAGCAGCGCGTCGACTGCCGTTAGCGCAGCGCCGACGGAGAAAGAATT

TCCGAAAACTCACTCCGACGTAAGCAGCGAGCAGCAACACCTGAAGGGCCAGACCGGCACTAACG
GTGGTAACGTGAAACTGTACACGACCGACCAGAGCACGTTCGATATTCTGCAGGACCTGGAATTCA
GCAGCGGCTCGCCGGGTAAAGAAACTAACGAAAGCCCGTGGCGTTCGGATCTGTTGATCGACGAA
AACTGTCTGCTGAGCCCGCTGGCCGGTGAGGACGATAGCTTTTTGCTGGAGGGTAATAGCAACGAG
GACTGCAAACCGCTGATCCTCCCGGACACCAAGCCGAAGATTAAGGATAATGGTGACTTGGTTCTG
AGCAGCCCGAGCAACGTGACCCTTCCGCAAGTCAAGACCGAGAAAGAGGACTTTATCGAGTTGTGC
ACCCCGGGTGTGATCAAGCAGGAGAAACTGGGCACCGTCTACTGCCAAGCGAGCTTCCCTGGTGC
CAATATCATCGGCAACAAGATGAGCGCGATCTCTGTTCACGGTGTATCTACGTCCGGCGGTCAAAT
GTACCACTATGACATGAATACCGCAAGCCTGTCGCAACAGCAAGACCAAAAGCCGATTTTCAACGT
GATCCCACCGATTCCGGTGGGTTCCGAGAATTGGAATCGTTGTCAGGGCTCTGGTGACGATAATCT
GACCTCACTGGGTACTCTGAACTTTCCGGGTCGCACCGTCTTTAGCAACGGTTATAGCTCACCGAG
CATGCGTCCTGATGTTAGCTCTCCGCCAAGCTCCAGCTCGACCGCGACCACCGGCCCACCGCCGA
AACTGGATAGCCTGGAGTTCATTGCTTCTAAATTGGCATAA

Translation of above C3-NTD single molecule construct:

MHHHHHHENLYFQGLNDIFEAQKIEWHEMGNDLGFPQQGQISLSSGETDLKLLEESIANLNR
STSVPENPKSSASTAVSAAPTEKEFPKTHSDVSSEQQHLKGQTGTNGGNVKLYTTDQSTFDILQDLE
FSSGSPGKETNESPWRSDLLIDENCLLSPLAGEDDSFLLEGNSNEDCKPLILPDTKPKIKDNGDLVLS
SPSNVTLPQVKTEKEDFIELCTPGVIKQEKLGTVYCQASFPGANIIGNKMSAISVHGVSTSGGQMYHY
DMNTASLSQQQDQKPIFNVIPPIPVGSENWNRCQGSGDDNLTSLGTLNFPGRTVFSNGYSSPSMRP
DVSSPPSSSSTATTGPPPKLDSLEFIASKLA

4.6.5.2 Mammalian Expression Constructs (pJ603 plasmid)

TSG101 full-length ORF:

ATGGCTGTGTCCGAATCTCAACTGAAAAAGATGGTGTCCAAGTATAAGTACCGCGACCTCACT
GTGCGCGAAACCGTGAATGTCATCACCCTGTACAAGGATTTGAAGCCCGTGCTCGACTCATATGTG

TTCAATGATGGATCGTCCCGGGAACTGATGAACCTGACCGGAACCATCCCAGTCCCATACCGGGGC

AACACCTACAACATCCCGATTTGTCTGTGGCTGCTGGACACTTACCCGTACAATCCGCCGATCTGC

TTCGTTAAACCGACCAGCTCCATGACTATCAAGACCGGCAAACATGTCGACGCGAATGGAAAGATC

TACCTCCCGTACTTGCACGAGTGGAAGCACCCGCAGTCAGATCTGCTTGGGTTGATCCAGGTCATG

ATCGTGGTGTTTGGAGATGAACCGCCGGTGTTCTCGCGCCCGATTAGCGCATCCTACCCTCCATAC

CAAGCCACCGGCCCACCAAACACTAGCTATATGCCCGGGATGCCCGGCGGGATATCCCCGTACCC

TTCCGGATACCCGCCTAATCCAAGCGGCTACCCTGGTTGCCCCTACCCACCTGGAGGACCCTACC

CTGCCACCACGAGCAGCCAGTACCCATCACAACCGCCAGTTACTACTGTGGGACCGTCGAGAGAT

GGTACTATCTCGGAGGACACGATCAGAGCTTCGCTCATCTCGGCGGTGTCAGACAAGCTTCGCTGG

AGGATGAAAGAGGAAATGGATAGGGCGCAGGCAGAGCTTAACGCCCTGAAGCGGACTGAAGAGGA

CCTCAAGAAAGGTCACCAGAAACTGGAAGAGATGGTCACGCGCCTGGACCAGGAAGTGGCCGAGG

TGGACAAAAACATTGAGCTGCTCAAGAAGAAGGACGAAGAACTGTCGAGCGCCCTCGAAAAGATGG

AAAACCAATCCGAGAACAATGATATTGATGAAGTGATCATCCCTACCGCCCCGTTGTACAAGCAGA

TCCTCAACCTCTACGCCGAGGAGAACGCCATCGAGGATACCATCTTCTACCTGGGAGAAGCACTTA

GACGCGGAGTCATCGACCTGGACGTGTTTCTGAAGCATGTCCGACTGCTCTCACGGAAGCAATTCC

AGCTGAGAGCACTGATGCAGAAGGCTCGGAAAACTGCGGGACTGTCGGACCTCTACTGA

Translation of TSG101:

MAVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYR

GNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMI

VVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTS

SQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQ

KLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEEN

AIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLSDLY

TSG101cc ORF:

ATGGCTTCGCTCATCTCGGCGGTGTCAGACAAGCTTCGCTGGAGGATGAAAGAGGAAATGGAT

AGGGCGCAGGCAGAGCTTAACGCCCTGAAGCGGACTGAAGAGGACCTCAAGAAAGGTCACCAGAA
ACTGGAAGAGATGGTCACGCGCCTGGACCAGGAAGTGGCCGAGGTGGACAAAAACATTGAGCTGC
TCAAGAAGAAGGACGAAGAACTGTCGAGCGCCCTCGAAAAGTGA

Translation of TSG101cc:

MASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNI
ELLKKKDEELSSALEK

TSG101Δcc ORF:

ATGGCTGTGTCCGAATCTCAACTGAAAAAGATGGTGTCCAAGTATAAGTACCGCGACCTCACT
GTGCGCGAAACCGTGAATGTCATCACCCTGTACAAGGATTTGAAGCCCGTGCTCGACTCATATGTG
TTCAATGATGGATCGTCCCGGGAACTGATGAACCTGACCGGAACCATCCCAGTCCCATACCGGGGC
AACACCTACAACATCCCGATTTGTCTGTGGCTGCTGGACACTTACCCGTACAATCCGCCGATCTGC
TTCGTTAAACCGACCAGCTCCATGACTATCAAGACCGGCAAACATGTCGACGCGAATGGAAAGATC
TACCTCCCGTACTTGCACGAGTGGAAGCACCCGCAGTCAGATCTGCTTGGGTTGATCCAGGTCATG
ATCGTGGTGTTTGGAGATGAACCGCCGGTGTTCTCGCGCCCGATTAGCGCATCCTACCCTCCATAC
CAAGCCACCGGCCCACCAAACACTAGCTATATGCCCGGGATGCCCGGCGGGATATCCCCGTACCC
TTCCGGATACCCGCCTAATCCAAGCGGCTACCCTGGTTGCCCCTACCCACCTGGAGGACCCTACC
CTGCCACCACGAGCAGCCAGTACCCATCACAACCGCCAGTTACTACTGTGGGACCGTCGAGAGAT
GGTACTATCTCGGAGGACACGATCAGAGGATCCGGCGGTGGATCTGGTGGATCTGGTGGATCTGG
TGGATCTGGTACCATGGAAAACCAATCCGAGAACAATGATATTGATGAAGTGATCATCCCTACCGC
CCCGTTGTACAAGCAGATCCTCAACCTCTACGCCGAGGAGAACGCCATCGAGGATACCATCTTCTA
CCTGGGAGAAGCACTTAGACGCGGAGTCATCGACCTGGACGTGTTTCTGAAGCATGTCCGACTGCT
CTCACGGAAGCAATTCCAGCTGAGAGCACTGATGCAGAAGGCTCGGAAAACTGCGGGACTGTCGG
ACCTCTACTGAAAAATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGT

Translation of TSG101Δcc:

MAVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYR
GNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMI
VVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTS
SQYPSQPPVTTVGPSRDGTISEDTIRGSGGGSGGSGGSGGSGTMENQSENNDIDEVIIPTAPLYKQIL
NLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLSDLY

TSG101–tdTom ORF (tdTom in lower-case letters):

ATGGCTGTGTCCGAATCTCAACTGAAAAAGATGGTGTCCAAGTATAAGTACCGCGACCTCACT
GTGCGCGAAACCGTGAATGTCATCACCCTGTACAAGGATTTGAAGCCCGTGCTCGACTCATATGTG
TTCAATGATGGATCGTCCCGGGAACTGATGAACCTGACCGGAACCATCCCAGTCCCATACCGGGGC
AACACCTACAACATCCCGATTTGTCTGTGGCTGCTGGACACTTACCCGTACAATCCGCCGATCTGC
TTCGTTAAACCGACCAGCTCCATGACTATCAAGACCGGCAAACATGTCGACGCGAATGGAAAGATC
TACCTCCCGTACTTGCACGAGTGGAAGCACCCGCAGTCAGATCTGCTTGGGTTGATCCAGGTCATG
ATCGTGGTGTTTGGAGATGAACCGCCGGTGTTCTCGCGCCCGATTAGCGCATCCTACCCTCCATAC
CAAGCCACCGGCCCACCAAACACTAGCTATATGCCCGGGATGCCCGGCGGGATATCCCCGTACCC
TTCCGGATACCCGCCTAATCCAAGCGGCTACCCTGGTTGCCCCTACCCACCTGGAGGACCCTACC
CTGCCACCACGAGCAGCCAGTACCCATCACAACCGCCAGTTACTACTGTGGGACCGTCGAGAGAT
GGTACTATCTCGGAGGACACGATCAGAGCTTCGCTCATCTCGGCGGTGTCAGACAAGCTTCGCTGG
AGGATGAAAGAGGAAATGGATAGGGCGCAGGCAGAGCTTAACGCCCTGAAGCGGACTGAAGAGGA
CCTCAAGAAAGGTCACCAGAAACTGGAAGAGATGGTCACGCGCCTGGACCAGGAAGTGGCCGAGG
TGGACAAAAACATTGAGCTGCTCAAGAAGAAGGACGAAGAACTGTCGAGCGCCCTCGAAAAGATGG
AAAACCAATCCGAGAACAATGATATTGATGAAGTGATCATCCCTACCGCCCCGTTGTACAAGCAGA
TCCTCAACCTCTACGCCGAGGAGAACGCCATCGAGGATACCATCTTCTACCTGGGAGAAGCACTTA
GACGCGGAGTCATCGACCTGGACGTGTTTCTGAAGCATGTCCGACTGCTCTCACGGAAGCAATTCC
AGCTGAGAGCACTGATGCAGAAGGCTCGGAAAACTGCGGGACTGTCGGACCTCTACGGATCCGGT
TCTGGCTCCGGCGGTTCAGGATCTGGTACCatggtgagcaagggcgaggaggtcatcaaagagttcatgcgcttcaa

ggtgcgcatggagggctccatgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagac
cgccaagctgaaggtgaccaagggcggccccctgcccttcgcctgggacatcctgtccccccagttcatgtacggctccaaggcgta
cgtgaagcaccccgccgacatccccgattacaagaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgagga
cggcggtctggtgaccgtgacccaggactcctccctgcaggacggcacgctgatctacaaggtgaagatgcgcggcaccaacttcc
cccccgacggccccgtaatgcagaagaagaccatgggctgggaggcctccaccgagcgcctgtacccccgcgacggcgtgctgaa
gggcgagatccaccaggccctgaagctgaaggacggcggccactacctggtggagttcaagaccatctacatggccaagaagccc
gtgcaactgcccggctactactacgtggacaccaagctggacatcacctcccacaacgaggactacaccatcgtggaacagtacgag
cgctccgagggccgccaccacctgttcctggggcatggcaccggcagcaccggcagcggcagctccggcaccgcctcctccgagg
acaacaacatggccgtcatcaaagagttcatgcgcttcaaggtgcgcatggagggctccatgaacggccacgagttcgagatcgag
ggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggcggccccctgcccttcgcctgg
gacatcctgtcccccccagttcatgtacggctccaaggcgtacgtgaagcaccccgccgacatccccgattacaagaagctgtccttcc
ccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggtctggtgaccgtgacccaggactcctccctgcaggacggc
acgctgatctacaaggtgaagatgcgcggcaccaacttcccccccgacggccccgtaatgcagaagaagaccatgggctgggagg
cctccaccgagcgcctgtacccccgcgacggcgtgctgaagggcgagatccaccaggccctgaagctgaaggacggcggccacta
cctggtggagttcaagaccatctacatggccaagaagcccgtgcaactgcccggctactactacgtggacaccaagctggacatcac
ctcccacaacgaggactacaccatcgtggaacagtacgagcgctccgagggccgccaccacctgttcctgtacggcatggacgagc
tgtacaagtaa

Translation of TSG101–tdTom:

MAVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYR
GNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMI
VVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTS
SQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQ
KLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEEN
AIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLSDLYGSGSGSGGSGSGT

MVSKGEEVIKEFMRFKVRMEGSMNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQF

MYGSKAYVKHPADIPDYKKLSFPEGFKWERVMNFEDGGLVTVTQDSSLQDGTLIYKVKMRGTNFPP

DGPVMQKKTMGWEASTERLYPRDGVLKGEIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDT

KLDITSHNEDYTIVEQYERSEGRHHLFLGHGTGSTGSGSSGTASSEDNNMAVIKEFMRFKVRMEGS

MNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYKKLSFP

EGFKWERVMNFEDGGLVTVTQDSSLQDGTLIYKVKMRGTNFPPDGPVMQKKTMGWEASTERLYPR

DGVLKGEIHQALKLKDGGHYLVEFKTIYMAKKPVQLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRH

HLFLYGMDELYK

4.6.6 Python Code

Code Used to Parse CD Data:

```
 #!/usr/bin/python
#Jordan White 2013


import sys
import numpy as np  #I am using numpy because it has useful matrix
    operations not intrinsic to python.


#Usage:  in a terminal window input:
# $  python3 <path to this script> <path to blank CD scan> <path to next
     data file>
# Input the values the script prompts you for.
```

```
# The output will be a csv file of the data along with a header to each
    column. Make sure to choose the location you want for the new file.
    Change np.savetxt's first object to change the save file location.
# Line 33 dictates where the data will be cut off. A cut off dynode
    voltage of 600.5 is pretty conservative. I have seen good data that
    had a dynode voltage closer to 700 or 800. User discretion advised.
    This script adds zeros at over-voltage lines for two reasons: it
    maintains the data structure (easier to code for) and I had trouble
    when trying to tell Python to read in an empty data value. *****
    IMPORTANT: Because this replaces over-voltaged lines with zeros, you
     should input your lowest concentration data (lowest dynode voltage)
     for this script. This gives you your maximum wavelength range.
    CD_Data_Parser_MRE_additional.py will then add additional data to
    the csv file.******  At the end you can delete the zeros by hand (
    excel) before importing to R or you can use slice notation in R.



def main():
#Inputs
    baseline = sys.argv[1]
    inputfile = sys.argv[2] #direct terminal input of file alongside the
        invocation of this script
    residue_concentration = float(input('What is the residue length
       minus one of your protein times its **Molar** concentration? '))
    hat = input('What header do you want for your data? ')
```

```
#
#
#Definition of data arrays and the header, "head", for the output
    raw = np.genfromtxt(inputfile, delimiter = '  ', skip_header = 19,
        skip_footer = 160, usecols = (0, 1, 2, 6))     #open my data,
        use two white spaces as the delimiter, skip the header and
        footer (footer changes depending on how many wavelengths are
        scanned), and only import the relevant columns. If you need to
        change skip_footer: trial and error can fix minor offsets or you
         can use bash ($tail -n-160) to get a sense of what lines are
        getting chopped by skip_footer.
    blank = np.genfromtxt(baseline, delimiter = '  ', skip_header = 19,
        skip_footer = 160, usecols = (0,1)) # Same as above except this
        is for the blank. Change skip_footer if the blank has a
        different wavelength range than the sample.
    data = np.array([])
    head = 'Wavelength,'+hat+','+ hat + '_error'
#
#
#Math Operations
    raw[:,1:2] = raw[:,1:2] - blank[:,1:2] #this does two things:
        subtracts the blank from my data and then replaces the entire
        data signal column with the blank subtracted signal
    zeroed = np.array([0,0,0]) #row to add to over-voltage lines
    for line in raw:
```

```
            if line[3:4] <= 600.5: #ignore data with a dynode signal > 600
                data = np.concatenate((data, line[0:3]), axis = 0) #add the
                    data to the matrix "data"
            else:
                data = np.concatenate((data, zeroed[0:3]), axis = 0) # adds
                    zeros to lines with maxed out voltage
        data.shape = (len(data)/3, 3) # reshapes the rows and columns
        MRE = data*[1, 1/(residue_concentration*1000),1/(
            residue_concentration*1000)]
        np.savetxt('/Users/jordanwhite/Documents/My_Research/Hilser_Lab/CD/
            DATA.csv', MRE, delimiter = ',', header = head, comments = '') #
            writes output held in "data"
        print('Done')
        print('')
if __name__ == "__main__":
        main()
```

Additional data was added using this:

```
#!/usr/bin/python
#Jordan White 2013

import sys
import numpy as np

#Usage:  in a terminal window input:
```

```
# $  python3 <path to this script> <path to blank CD scan> <path to csv
    file of data already parsed> <path to next data file>
# input the values the script prompts you for
# the output will be a csv file of the previously parsed data plus the
    next dataset. Make sure to choose the location you want for the new
    file. Change np.savetxt's first object as required.


def main():
#Input
    baseline = sys.argv[1]
    previous = sys.argv[2]
    inputfile = sys.argv[3] #direct terminal input of file alongside the
        invocation of this script
    residue_concentration = float(input('What␣is␣the␣residue␣length␣of␣
        your␣protein␣times␣its␣**Molar**␣concentration?␣'))
    hat = input('What␣header␣do␣you␣want␣for␣your␣data?␣')
#
#Definition of arrays
    raw = np.genfromtxt(inputfile, delimiter = '␣␣', skip_header = 19,
        skip_footer = 160, usecols = (1, 2, 5))     #open my data, skip
        the header and footer, only import the relevant columns
    blank = np.genfromtxt(baseline, delimiter = '␣␣', skip_header = 19,
        skip_footer = 160, usecols = (0,1)) #same as above except for
        blank, change skip_footer if the blank has a different
        wavelength range than the sample
```

```
    dataset = np.genfromtxt(previous, delimiter = ',', skip_header = 1)
    dataset_csv = csv.reader(open(previous))
    data = np.array([])
    dataset_header = ','.join(next(dataset_csv)) #reads in first line of
        previous data set (the header) and uses comma separation
    head = dataset_header +','+ hat + ','+ hat + '_error' #appends
        previous and current headers
#
#Math
    raw[:,0:1] = raw[:,0:1] - blank[:,1:2] #this does two things:
        subtracts the blank from my data and then replaces the entire
        data signal column with the blank subtracted signal
    zeroed = np.array([0,0])
    for line in raw:
        if line[2:3] <= 600.5: #ignore data with a dynode signal > 600
            data = np.concatenate((data, line[0:2]), axis = 0) #add the
                data to the matrix "data"
        else:
            data = np.concatenate((data, zeroed[0:2]), axis = 0)
    data.shape = (len(data)/2, 2)
    MRE = data*[1/(residue_concentration*1000),1/(residue_concentration
        *1000)]
    output = np.concatenate((dataset, MRE), axis = 1) # appends previous
        dataset to the current dataset
    np.savetxt('/Users/jordanwhite/Documents/My_Research/Hilser_Lab/CD/
```

```
        DATA.csv', output, delimiter = ',', header = head, comments = ''
        ) #writes output held in "data"
    print('Done')
    print('')
#
if __name__ == "__main__":
    main()
```

When mixing two proteins together, the difference spectrum was determined using the following:

```
#!/usr/bin/python


import sys
import csv
import numpy as np


#This will determine the difference between measured and expected CD
    spectra of protein mixtures. The input requires that there already
    be a DATA file, and the output is in MRE.
#Usage:  in a terminal window input:
# $  python3 <path to this script> <path to blank CD scan> <path to csv
    file of data already parsed> <path to next data file>
# input the values the script prompts you for
# the output will be a csv file of the previously parsed data plus the
    next dataset. Make sure to choose the location you want for the new
```

```
    file. Change np.savetxt's first object as required.


def main():
    baseline = sys.argv[1]
    previous = sys.argv[2]
    inputfile = sys.argv[3] #direct terminal input of file alongside the
        invocation of this script
    proteinA = sys.argv[4] #CD spectrum of one of the proteins by itself
        ****This requires that this CD spectrum is at the same
       concentration as the protein measured**** Otherwise, there is a
       linear concentration factor that multiplies the mdeg measured,
       assuming no aggregation.
    proteinB = sys.argv[5] #CD spectrum of the other protein by itself.
       Same caveat as above.
    residue_concentration = float(input('What␣is␣the␣residue␣length␣of␣
       your␣protein␣times␣its␣**Molar**␣concentration?␣')) #[ProteinA
       ]*(length of A - 1)+[ProteinB]*(length of B - 1)
    hat = input('What␣header␣do␣you␣want␣for␣your␣data?␣')
    raw = np.genfromtxt(inputfile, delimiter = '␣␣', skip_header = 19,
       skip_footer = 160, usecols = (1, 2, 5))     #open my data, skip
       the header and footer, only import the relevant columns
    blank = np.genfromtxt(baseline, delimiter = '␣␣', skip_header = 19,
       skip_footer = 165, usecols = (0,1)) #same as above except for
       blank, change skip_footer if the blank has a different
       wavelength range than the sample
```

```
    protA = np.genfromtxt(proteinA, delimiter = '  ', skip_header = 19,
        skip_footer = 160, usecols = (0,1))
    protB = np.genfromtxt(proteinB, delimiter = '  ', skip_header = 19,
        skip_footer = 160, usecols = (0,1))
    dataset = np.genfromtxt(previous, delimiter = ',', skip_header = 1)
    dataset_csv = csv.reader(open(previous))
    data = np.array([])
#
    dataset_header = ','.join(next(dataset_csv))
    head = dataset_header +','+ hat + ','+ hat + '_error'
#
#
    raw[:,0:1] = raw[:,0:1] + blank[:,1:2] - protA[:,1:2] - protB[:,1:2]
        #this does two things:  subtracts the blank from my data and
        then replaces the entire data signal column with the blank
        subtracted signal **Note:  The raw data of protein A and B
        contain the buffer signal. Thus, when the raw protein A and B
        signals are subtracted from the combined signal, buffer is
        subtracted twice; therefore, one buffer signal must be added
        back.**
    zeroed = np.array([0,0])
    for line in raw:
        if line[2:3] <= 600.5: #ignore data with a dynode signal > 600
            data = np.concatenate((data, line[0:2]), axis = 0) #add the
                data to the matrix "data"
```

```
            else:
                data = np.concatenate((data, zeroed[0:2]), axis = 0)
        data.shape = (len(data)/2, 2)
        MRE = data*[1/(residue_concentration*1000),1/(residue_concentration
            *1000)]
        output = np.concatenate((dataset, MRE), axis = 1)
        np.savetxt('/Users/jordanwhite/Documents/My_Research/Hilser_Lab/CD/
            DATA.csv', output, delimiter = ',', header = head, comments = ''
            ) #writes output held in "data"
        print('I\'m␣done␣bitch')
        print('')
#
if __name__ == "__main__":
    main()
```

Chapter 5

Identification of potential GR drug-leads

5.1 Introduction

Glucocorticoid receptor (GR) is a desirable drug target because of its connection to a wide variety of disease states, including: suicide/depression, both activation and suppression of the immune system, prostate and breast

cancer, and cardiovascular dysfunction [19, 145, 187, 275, 308, 456]. All current glucocorticoid drugs target the ligand binding domain (see Figure 5.1). This is a vulnerability in our drug inventories because the ligand binding domain (LBD) is disposable to GR function. Removal of the LBD causes GR to constitutively localize in the nucleus of cells where it binds DNA and alters the expression of a wide variety of genes [260].

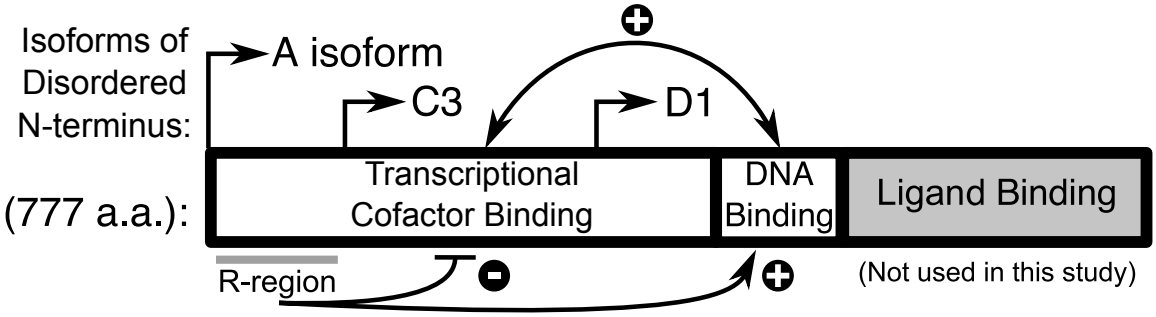


Figure 5.1 – The Domain Organization of GR
The disordered N-terminal domain (NTD) has eight translational isoforms (three are shown). The NTD binds transcriptional cofactor proteins close to the C3 isoform start site [260]. Because the D1 isoform is missing the region that is critical for transcriptional activation, it (or the DBD alone) can only act as a repressor. The DNA-binding domain (DBD) binds DNA response elements as a dimer or monomer (glucocorticoid response elements, GRE's). The ligand-binding domain (LBD) binds steroids and controls GR's nuclear localization, but it is dispensable for GR's core functions. The themodynamic interactions within GR were characterized by a previous student (Dr. Li), and are shown as arrows with "+" for stabilizing and "-" for destabilizing interactions. Note that the Regulatory-region (R-region) of the NTD can have both negative and positive effects on transcriptional activity [242, 243]. The domains are scaled to their respective lengths.

For a disease that evolves, namely cancer, it could mutate GR such that the ligand binding domain would no longer bind drugs. Indeed, the natural GR gene (NR3C1) already produces at least two splice isoforms without functional ligand binding domains [350], and it has been shown that several hematological cancers can upregulate the production of one such isoform, called GR-P [86]. The P

isoform cannot bind glucocorticoid drugs and this is one of several ways in which these cancers can become refractory to glucocorticoid treatment.

If we discover drugs that bind the GR DNA-binding domain (DBD), then we will gain a fundamentally different glucocorticoid drug. Not only would such a drug be resistant to loss of the ligand binding domain, but it would also open up a whole new field of GR drug design. By directly targeting the DBD, we may gain more control over what genes GR targets for transcriptional regulation.

Previously, a student in the Hilser lab, Dr. Jing Li, determined that there was a Regulatory (R) region within the GR N-Terminal Domain (NTD) [242]. The R-region destabilizes the folded state of the NTD, in particular the cofactor binding region close to the C3-isoform start site (Figure 5.1). However, the R-region also stabilizes the DBD. This stabilization leads to enhanced DNA-binding which can stabilize the cofactor binding region of the NTD [218]. These conflicting roles of the R-region produce a bistable system where small perturbations can lead to large changes in transcriptional activity. To determine the residues that mediate the DBD to R-region interaction, Dr. Li mutated several conserved DBD residues and found some that were necessary for the DBD to R-region interaction. Strikingly, the residues form a surface on the DBD structure. The Hilser lab proposed this surface was an allosteric site [243], and my collaborator, Dr. E. Brad Thompson, suggested that I dock small molecules onto this DBD site, in the hopes of finding a drug lead.

The effect such a drug would have on GR activity can be predicted because of the thermodynamic model that Dr. Li built using the Hilser lab's ensemble allosteric model [243, 289]. Such a drug could either block the NTD to DBD interaction or mimic the interaction. In both cases, the probability of the A isoform of GR being active would go from about 10% to 30%. Increasing the activity of the A isoform could push leukemias into apoptosis [407] (given other chemotherapy drugs) or it could give us a targeted therapy against some other disease: the A isoform is expressed to a different degree in different tissues [260].

In the proceeding chapter, I detail my verification of the allosteric site and my attempts to find drug-leads using the DOCK 6.6 software.

5.2 Methods

5.2.1 Luciferase assays

U2OS cells were grown in a 5% CO_2 atmosphere at 37º C and transfected at 90% confluence using XtremeGENE HP (Roche). The luciferase plasmids expressed either *Gaussia* luciferase or *Cypridina* luciferase (both in Mini-TK 2 vector of NEB). GR normally binds specific DNA sites, GR response elements (GRE), as a dimer. A GRE can be split in half, yielding a single GR binding site, without the convoluting effect of dimerization. To prevent dimerization effects in the mutational analysis, four GRE half-sites were cloned into the *Gaussia* promoter.

As any given plasmid was titrated, an empty pJ603 backbone was used to balance the total amount of DNA transfected.

Two luciferase assays are presented in this section. In both assays, *Gaussia* and *Cypridina* luciferases and a C3-DBD expression plasmid (pJ603, DNA2.0) were transfected at constant amounts (40 ng, 40 ng, and 3 ng, respectively). The first experiment was to determine the residues of the DBD required for the interaction of the DBD with the R-region. In this case, the cells were transfected with up to 16 ng of a competitor, DBD expressing plasmid (WT, C431Y, V435A, or L436A). As a control, some cells were transfected with 16 ng of the DBD plasmid and no C3-DBD plasmid. The second experiment was to confirm the DBD to R-region interaction by expressing the R-region (up to 12 ng of plasmid) in the presence of a C3-DBD construct (WT or triple mutant: C431Y, V435A, L436A).

The medium of the cells was collected two days after transfection and luciferase activity was measured using kits from NEB and a Berthold luminometer. The data were baselined for *Gaussia* and *Cypridina* signal in untransfected wells. The averages presented here are from four to six replicates ± 95% confidence intervals. T-tests were done in R 3.1.2, with a Benjamini-Hochberg correction applied at an alpha of 1% [42].

5.2.2 Cloning of constructs

A GeneBlock was synthesized by IDT to contain the GR R-region (1-97 a.a.), a GSx2 linker, the GR nuclear localization sequence (488-505 a.a.), a GSx3 linker, and the FLAG tag (DYKDDDDK). This GeneBlock was restriction digested with NheI and XhoI, then inserted into the pJ603 vector (DNA2.0) for mammalian expression. Clones were sequenced using the Sanger method (Genewiz).

Other plasmids used here were cloned by Dr. Jing Li during her thesis research. The R-DBD constructs have the R-region of GR (a.a. 1-97) attached to the DBD via an 11a.a. linker (GTGGSGGSGGS), and the related control construct lacks the R-region (see Figure 5.2, part A).

5.2.3 Immunofluorescence

U2OS cells were grown in 6-well plates and transfected with 100 ng of R-NLS-FLAG expressing plasmid and 900 ng of salmon sperm DNA. Cells were grown on German, circular cover slips for two days before being washed with PBS+2mM MgCl2 (PBSM) and fixed with 4% PFA. After 10 minutes, the cells were washed with PBSM three times and quenched with 50 mM NH4Cl for 10 minutes. Cells were permeabilized with the following solution (BTPAD) for 30 minutes: PBS+0.1% TX-100+1% BSA+0.5 µg/mL DAPI. Anti-FLAG antibody (*Mus* M2 from Sigma, gift of Kuruvilla lab at Johns Hopkins) was used at 1:1000

in BTPAD to label the R-NLS-FLAG construct for one hour. After washing three times with PBSM, secondary antibody at 1:400 was incubated for 30 minutes (Cell Signaling Technologies, goat anti-mus Alexa 594) and washed off with PBSM three times. Slides were mounted with fluoromount (Sigma) and dried overnight. Images were taken using a Zeiss LSM 700 confocal microscope.

5.2.4 Docking

Ligands for docking were obtained from the ZINC clean-leads database, which is curated to be filled with non-toxic and drug-like molecules that mostly follow Lipinski's Rules of Five [174, 250]. After docking the clean-leads, Prof. Freire communicated to me that most corporations use fragment screening because the downstream optimization is easier and more reliable. After finding pockets that bind ligand fragments, one can link the fragments and form a full lead-like molecule. I have also docked the ZINC frags-now database (see appendix), but all the analyses presented here are for the clean-leads.

The DOCK 6.6 suite of programs [220, 221] was used for docking of ligands to the known structure of the GR DBD of rat [263] (the rat and human DBD's have 100% sequence identity—a fact that may not have been known when the crystal was solved in 1991). The DBDs of other steroid hormone receptors were prepared for docking via UCSF's Chimera [99, 317]. Docking proceeded first by grid scoring, then by AMBER scoring of the top hits [159, 427, 428] (see appendix

for input scripts). All grid computations were done using a cluster of four AMD Opteron processors running CentOS, and following grid scoring, the hits were screened manually for any unrealistic binding modes. All AMBER scoring was done using a 2.6 GHz Intel i7 processor core of a 2012 Macbook Pro running OSX 10.9.4. Jobs were submitted to the CentOS cluster using HTCondor [402]. For AMBER docking, the DBD of GR was prepped using H++ version 3.1 at a pH of 7.2 [14], and the docking proceeded in a parallel fashion using MPICH2 (http://www.mpich.org). After AMBER docking, the top hits were run through SciFinder and one hit (ZINC 06002995) was found to already act as a drug for carbonic anhydrase.

Dendrograms and heat maps were created using the ChemmineR package of R [70]. The top hits were converted into their binary, chemical fingerprints and then clustered using complete linkage. Quantitative Structure Activity Relationship (QSAR) analysis was attempted, but failed cross-validation checks (using the Open3DQSAR package).

5.2.5 Alignment of the steroid hormone receptors (SHR's)

Chimera was used to align the structures and sequences of GR and five other SHRs. Sequence alignment was done with a BLOSSUM80 matrix because the DBDs of various SHRs are highly related. The following is a list of the SHRs used along with citations for the structures: glucocorticoid, progesterone,

mineralocorticoid, androgen, estrogen-1, and an ancestral resurrected receptor [164, 263, 276, 344, 361, 365].

5.3 Results

5.3.1 Verification of the allosteric site

The allosteric site that Dr. Jing Li discovered is made of residues C431, V435, and L436 in the DBD of GR. To discover this site, Dr. Li attached the DBD to the R-region via a flexible, eleven amino acid linker (R-11aa-DBD). She then mutated several conserved residues on the DBD (the mutation strategy is described in her thesis), and used the same construct minus the R-region as a control (11aa-DBD). Each of these constructs could bind the promoter of the *Gaussia* luciferase plasmid, but they could only block transcription. Thus, transfection of either the R-11aa-DBD or 11aa-DBD construct lowered the transcriptional activity of the active, C3-DBD construct through competitive inhibition (Figure 5.2). By transfecting increasing amounts of the mutant DBD constructs, Dr. Li quantified the inhibition strength of each construct—called the IC50 here, that is the amount of DBD plasmid that brings the assay to 50% of its initial signal.

In this assay, the R-11aa-DBD construct inhibits the C3-DBD construct more readily than DBD alone because the R-region stabilizes the attached DBD, increasing its DNA-binding affinity. Mutations that disrupt the R to DBD coupling

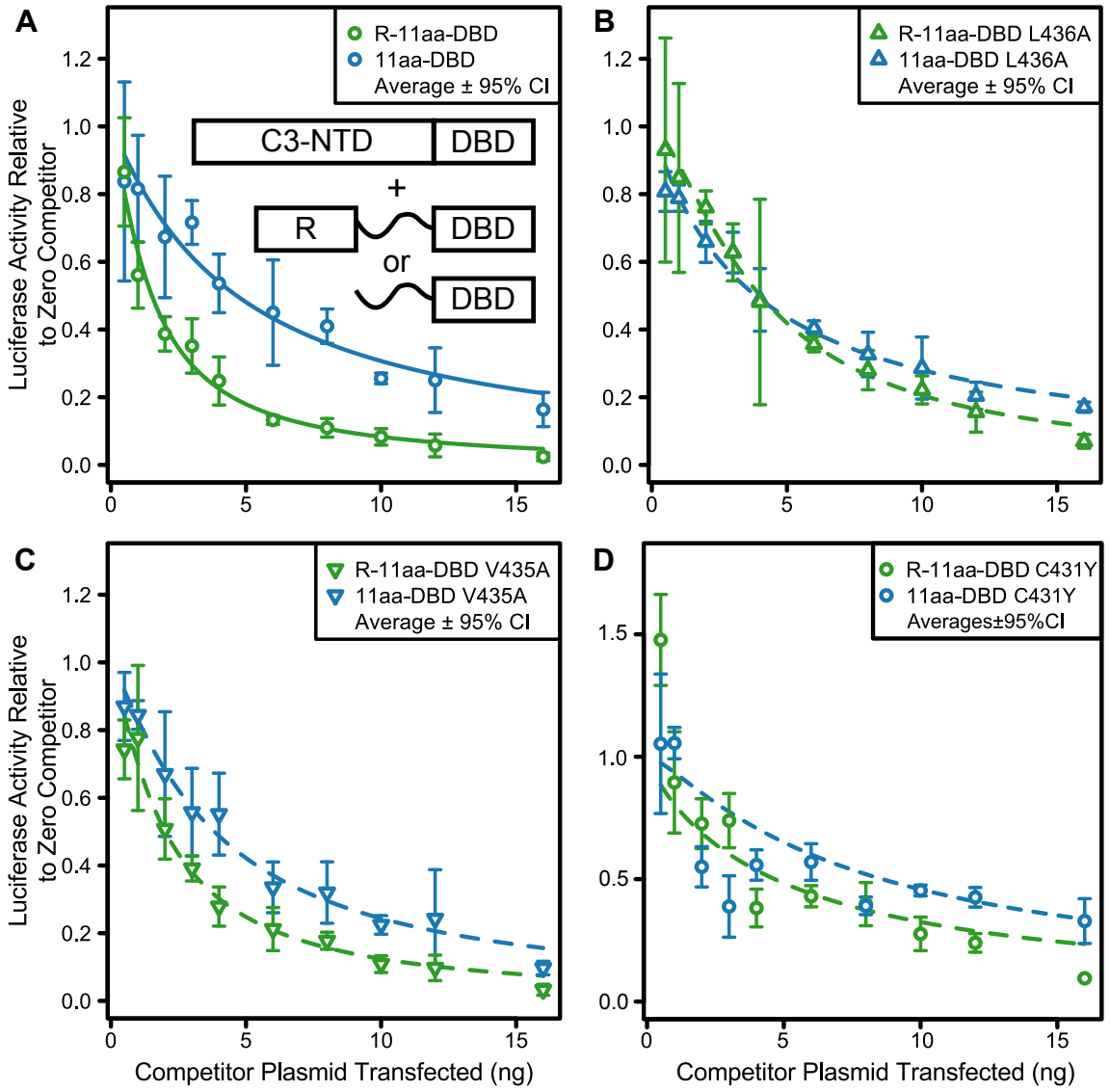


Figure 5.2 – **Verification of the DBD Allosteric Site**
A) The C3-DBD construct is inhibited by increasing amounts of either an R-DBD or DBD expression plasmid. This is Dr. Li's wild type data (R-DBD IC50 = 1.51±0.09 ng; DBD IC50 = 4.65±0.42 ng). The inset diagrams represent the constructs used here. **B)** Dr. Li's data for L436A, reanalysed (R-DBD IC50 = 4.01±0.35 ng; DBD IC50 = 3.85±0.18 ng). **C)** Dr. Li's data for V435A, reanalysed (R-DBD IC50 = 2.01±0.13 ng; DBD IC50 = 3.83±0.25 ng). **D)** The author's replication of the C431Y data (R-DBD IC50 = 4.81±0.58 ng; DBD IC50 = 5.88±0.89 ng).

will cause the R-11aa-DBD construct to have the same IC50 as the control DBD construct. Mutations that disrupt the DBD will cause both IC50's to shift, and some mutations will have no effect. As a first step in verifying Dr. Li's allosteric site, I reanalyzed her data. There are two equations to be considered, and the first describes the luciferase activity relative to zero competitor plasmid (y-axis of Figure 5.2):

$$\text{Activity relative to zero competitor plasmid} = \frac{\mathcal{Y} - \mathcal{B}}{\mathcal{Z} - \mathcal{B}} \tag{5.1}$$

Where $\mathcal{Y}$ is a sample's luciferase ratio, $\mathcal{B}$ is the baseline activity observed after transfecting 16 ng of RDBD or DBD construct alone, and $\mathcal{Z}$ is the luciferase ratio given 0 ng of competitor plasmid (C3DBD construct alone). The data were then fit to the following equation, describing the IC50:

$$y = 1 - \frac{1}{1 + (\frac{\mathcal{IC}50}{\mathcal{X}})^{p}} \tag{5.2}$$

where $\mathcal{X}$ is the ng of competitor plasmid and p is an arbitrary fitting parameter that accounts for the cooperativity in the curves.

Much of Figure 5.2 (parts a-c) is Dr. Li's data with my analysis. One mutant, C431Y yielded a borderline positive result upon reanalysis; thus, I recollected the data myself and present my own data in Figure 5.2d. I am omitting data for five other mutations because they are not involved in the allosteric site (L422,

S425, V449, L482, T493). My conclusions were unchanged by the reanalysis, thus I moved to an alternative test of the allosteric site.

I realized that instead of just inhibiting the C3-DBD construct, I could also activate it using the R-region of GR. The R-region would bind and stabilize the DBD of the C3 isoform, thereby increasing the transcriptional activity observed. Because the R-region would be dislocated from its natural position next to the C3-NTD, the R-region would likely be unable to destabilize the C3-NTD (large entropic penalty). If the allosteric site is truly the point of R and DBD interaction, then mutating it will ablate the interaction (C431Y, V435A, L436A). For this experiment to work, I designed an R-region construct that translocates into the nucleus on its own (Figure 5.3 and see Docking Methods section). As seen in Figure 5.4, the R-region activates the wild type C3 isoform of GR better than the triple mutant of GR.

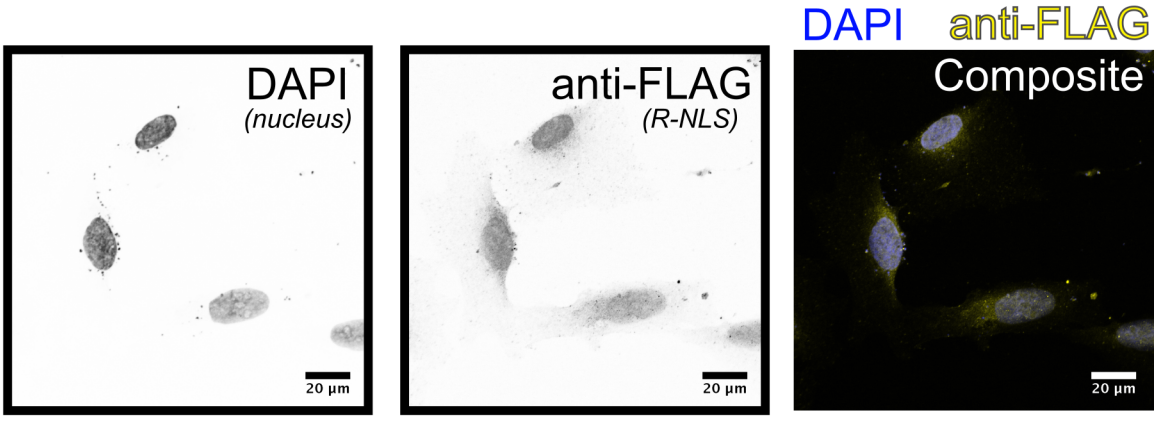


Figure 5.3 – Confocal Immunofluorescence of the R-NLS-FLAG Construct
Our R-region construct largely localizes to the nucleus (quantification = 48% nuclear ±5% given a 95% confidence interval, n = 7 cells). From left to right, the first panel is DAPI staining the DNA of the U2OS cells (signal inverted), an anti-FLAG antibody labeling the R-region construct (signal inverted), and a composite image with DAPI colored blue and the FLAG labeling colored yellow.

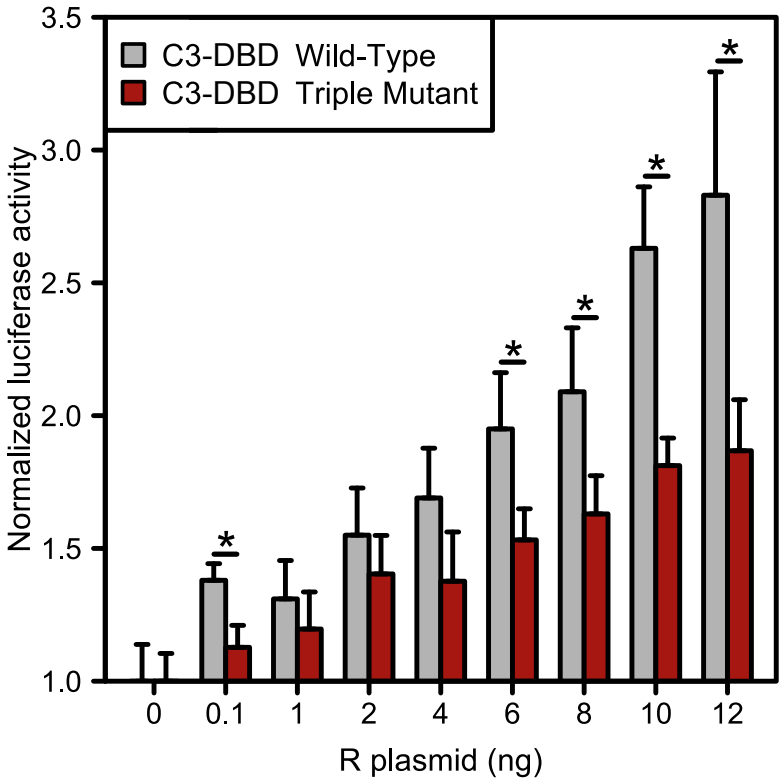


Figure 5.4 – **Luciferase Assay of C3-DBD with the R-NLS-FLAG Construct**
Increasing amounts of the R plasmid cause increasing activity of the C3 isoform of GR. However, the wild type C3 isoform is activated more than the DBD mutant (C431Y, V435A, L436A). Bars are means ± 95% confidence intervals and asterisks indicate statistically significant differences between bars with the same x-axis value (t-tests with Benjamini-Hochberg correction, $P < 0.01$).

5.3.2 In silico drug-lead screening

To find drugs that bind the DBD of GR, I used *in silico* drug-lead screening. This is a method whereby millions of small molecules are "docked" to the known structure of your favorite protein. The docking process is described in greater detail in the methods of this chapter. Briefly, the allosteric site of GR's DBD was targeted for docking and molecular force fields were used to calculate a docking

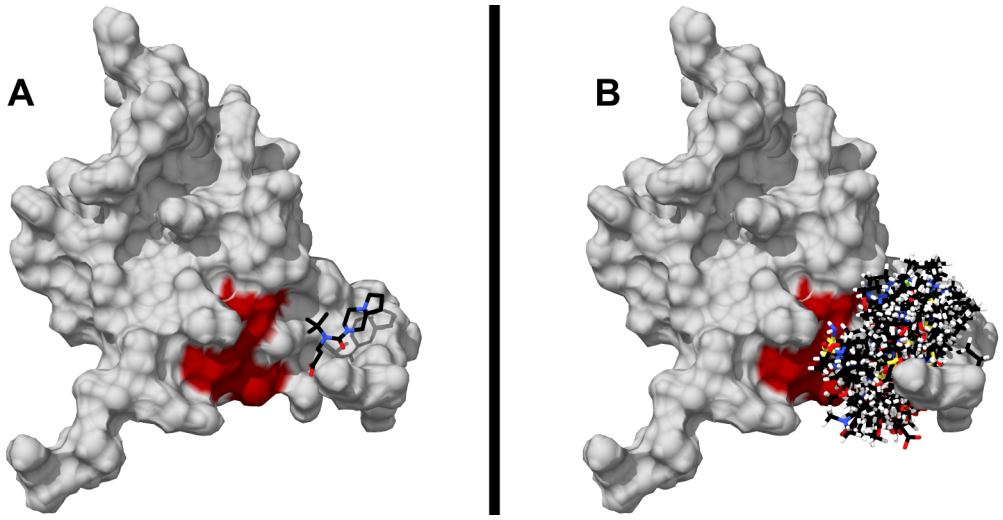


Figure 5.5 – The Allosteric Surface of the GR DBD
In both images, the DBD is colored as a gray surface with the three allosteric residues colored red. Note that they form a contiguous surface. **A)** The highest scoring hit (ZINC ID 94463428) docked to the DBD of GR. **B)** The top 105 hits docked to the DBD of GR. Note that the hits generally cluster around the allosteric site or next to it.

score for each drug-lead. The more negative the score, the better. See Figure 5.5 for images of the clean drug-leads docked to GR's DBD. Table 5.2 and Table 5.3 in the appendix list the top clean leads and frags-now leads, respectively, and Figure 5.6 has the structures of some of the top hits. The clean-lead results have been analyzed to the greatest degree and are presented in the main section here. The frags-now top hits are presented as an addendum in the appendix.

Clustering the top-hits of a drug screen can reveal recurring motifs and potentially aid future design of drugs. Figure 5.7 shows the clustering of the top 105 hits by chemical similarity. This revealed six major clusters with diverging chemical properties. There are a few commonalities among the clusters: An enrichment of aromatic and sulfa or carboxyl moieties (see Table 5.1).

Figure 5.6 – Structures of Top Hits

Cluster 1Aa, ZINC94463428, Score = -89

Cluster 1Ab, ZINC20534671, -47

Cluster 2Ba, ZINC77723236, -45

Cluster 1Aa, ZINC92625407, -71

Cluster 1Ab, ZINC91675613, -33

Cluster 2Ba, ZINC93091684, -43

Cluster 1B, ZINC92451923, -80

Cluster 2A, ZINC06002995, -58

Cluster 2Bb, ZINC94221043, -52

Cluster 1B, ZINC94221031, -66

Cluster 2A, ZINC92451895, -37

Cluster 2Bb, ZINC92263294, -42

Aromatic groups contribute to hydrophobic and pi-stacking interactions. However, the likely interaction in this case is hydrophobic. The drug-leads are not forming pi-stacking interactions with the DBD because the closest aromatic amino acid is tyrosine-433 and it faces away on its β-strand. The enrichment in sulfa and carboxyl moieties is likely related to the large number of basic residues in the GR DBD. The allosteric site of the DBD is next to two lysines and two arginines, and the drug-leads often docked with negatively charged moieties facing these positively charged residues of the DBD. Future development of these drug-leads should focus on optimizing these binding characteristics.

Table 5.1 – Chemical Make Up of the Drug-Lead Clusters
The first column is the cluster from Figure 5.7 (see the top dendrogram). Subsequent columns declare the percentage of molecules in a cluster that have a given chemical moiety. Within a given column, some of the clusters appear enriched with certain chemical groups and are highlighted for the reader.

Cluster	Aromatic	Urea	Amide	Sulfa	Carboxyl	Halogen
1Aa	21%	29%	14%	57%	100%	7%
1Ab	88%	13%	17%	88%	75%	4%
1B	50%	10%	0%	100%	100%	0%
2A	100%	20%	20%	80%	60%	0%
2Ba	100%	18%	12%	94%	59%	41%
2Bb	100%	28%	20%	100%	40%	8%

Typically, when one does *in silico* drug screening, there should be a positive control that is known to bind the target protein. The simulations shown here are completely *de novo*, thus I did not have the luxury of a positive control. To test the specificity of the docked ligands, I subsequently docked my top hits to the DBD's of five other steroid hormone receptors (SHR's). These SHR's

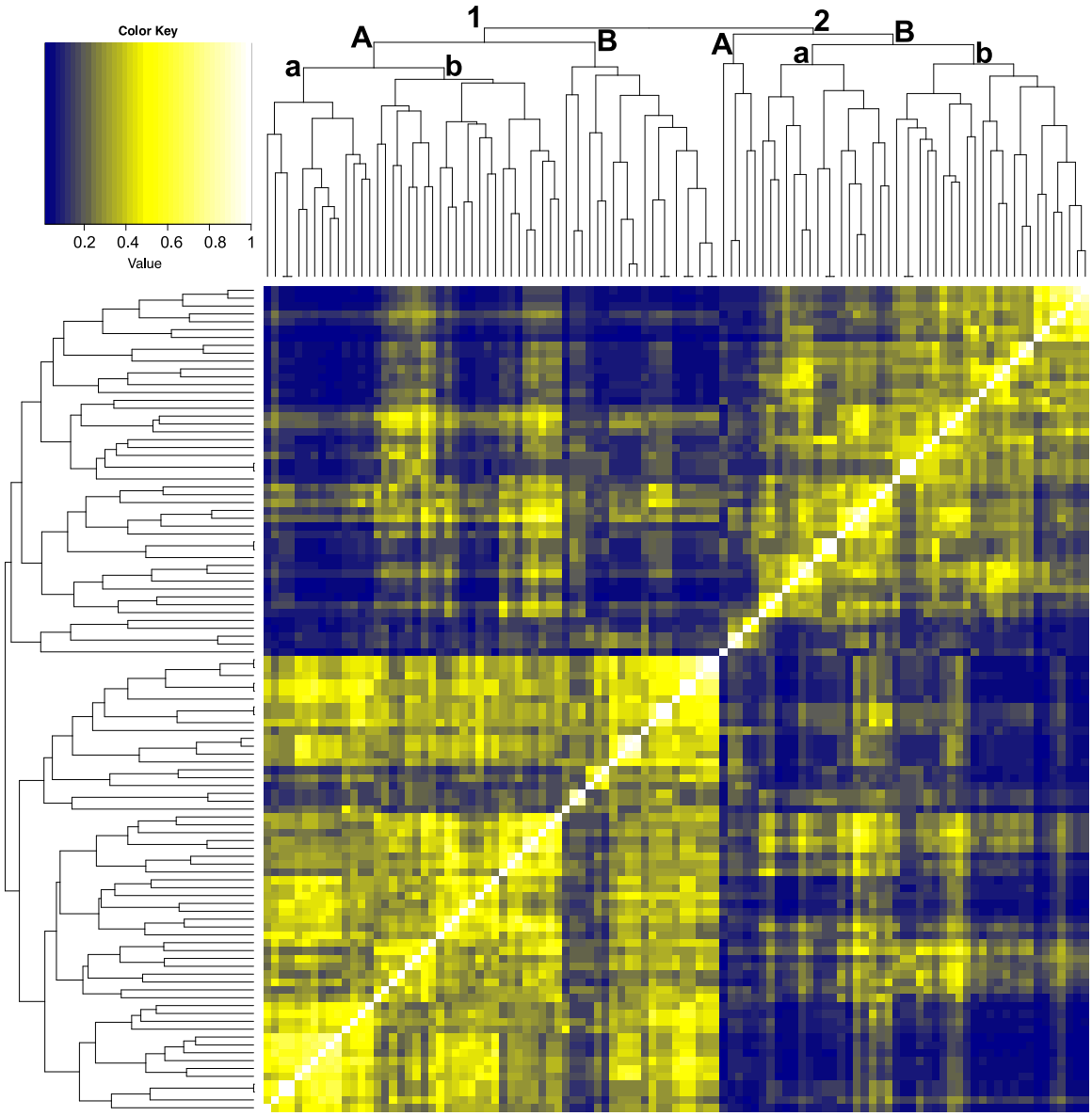
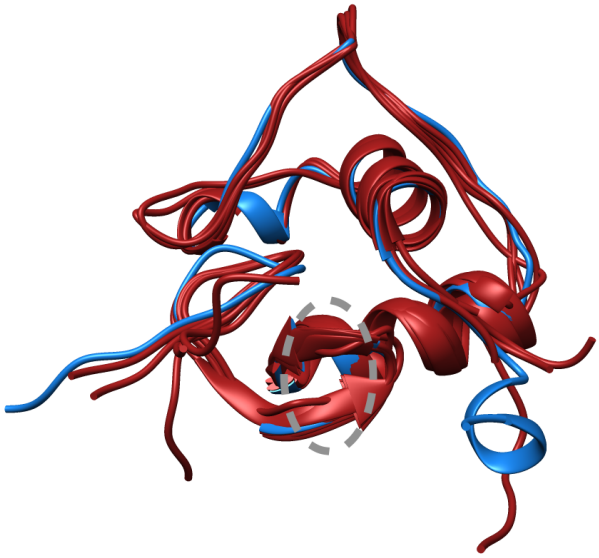


Figure 5.7 – Clustering of the Top 105 Docked Molecules
The top 105 hits cluster largely into two groups of chemical similarity. The names of prominent clusters are on the top dendrogram. See Table 5.1 for a break-down of the chemical moieties that are common in each cluster. The colors in this graphic depict chemical similarity, with blue being the least similar, yellow being similar, and white being the same or very similar. There are blocks of white near the diagonal because some of the hits are enantiomers.

are all closely related to GR (53—89% identity between the various DBD's), and their structures are strikingly similar (aligned backbone RMSD of between 0.540—0.768 Å). Figure 5.8 shows the structural alignment of the six DBD's and the sequence alignment around the allosteric site of GR.



GR:	G	C	HYG	VL	T
PR:	G	C	HYG	VL	T
MR:	G	C	HYG	VV	T
AR:	G	A	HYG	AL	T
ER1:	G	Y	HYG	VW	S
AncR:	G	F	HYG	VV	S

Figure 5.8 – Alignment of other Steroid Hormone Receptors to GR
Left: A structural alignment of GR (blue) and five other steroid hormone receptors (SHR's in red). The SHR's presented are as follows: progesterone (**PR**), mineralocorticoid (**MR**), androgen (**AR**), estrogen 1 (**ER1**), and an ancestral resurrected receptor (**AncR**). The allosteric site of GR is circled with a dashed, gray circle. **Right:** The sequence alignment around the allosteric site of GR. The dashed boxes indicate the relative position of the allosteric site residues (C…VL in GR). Blue letters indicate differences relative to the GR sequence.

In order to compare the docking scores, I calculated preference ratios, defined as the average docking score of a given receptor, divided by the average docking score of GR. A preference ratio greater than one indicates that the alternative receptor docks to the GR drug-leads better than GR itself. Based on the sequence alignment, I expected progesterone receptor to dock well to my top drug-leads, and it does indeed have the highest preference ratio (1.29). The other receptor

preference ratios are as follows: ancestral = 0.42, mineralocorticoid = 0.65, androgen = 0.36, and estrogen-1 = 0.68 (see Figure 5.9 and Figure 5.10 for comparison of raw docking scores).

The high preference ratio of progesterone receptor suggests that it may bind these drug-leads better than GR does. There are two mitigating pieces of information. First, the drug-leads dock to progesterone receptor in a manner that is less specific to the allosteric site and occluded by DNA-binding (see right side of Figure 5.10). Second, it is sometimes desirable to target both progesterone and glucocorticoid receptors in the clinic; thus, off-target effects on progesterone receptor may be desirable (personal communication with my collaborator Dr. E. Brad Thompson).

5.4 Conclusions

For decades, the ligand binding domain has been our only drug target for glucocorticoid receptor. This produced great medical advances, but was inherently limiting. In this chapter (and in Dr. Li's thesis), I describe an alternative drug target on GR—an allosteric site on the DBD. Over a million small molecules were docked to the DBD site, and the top hits found here will inform future drug discovery efforts. The top hits yielded information on the chemical composition we should expect of a GR DBD drug. Future work in the Hilser lab will identify

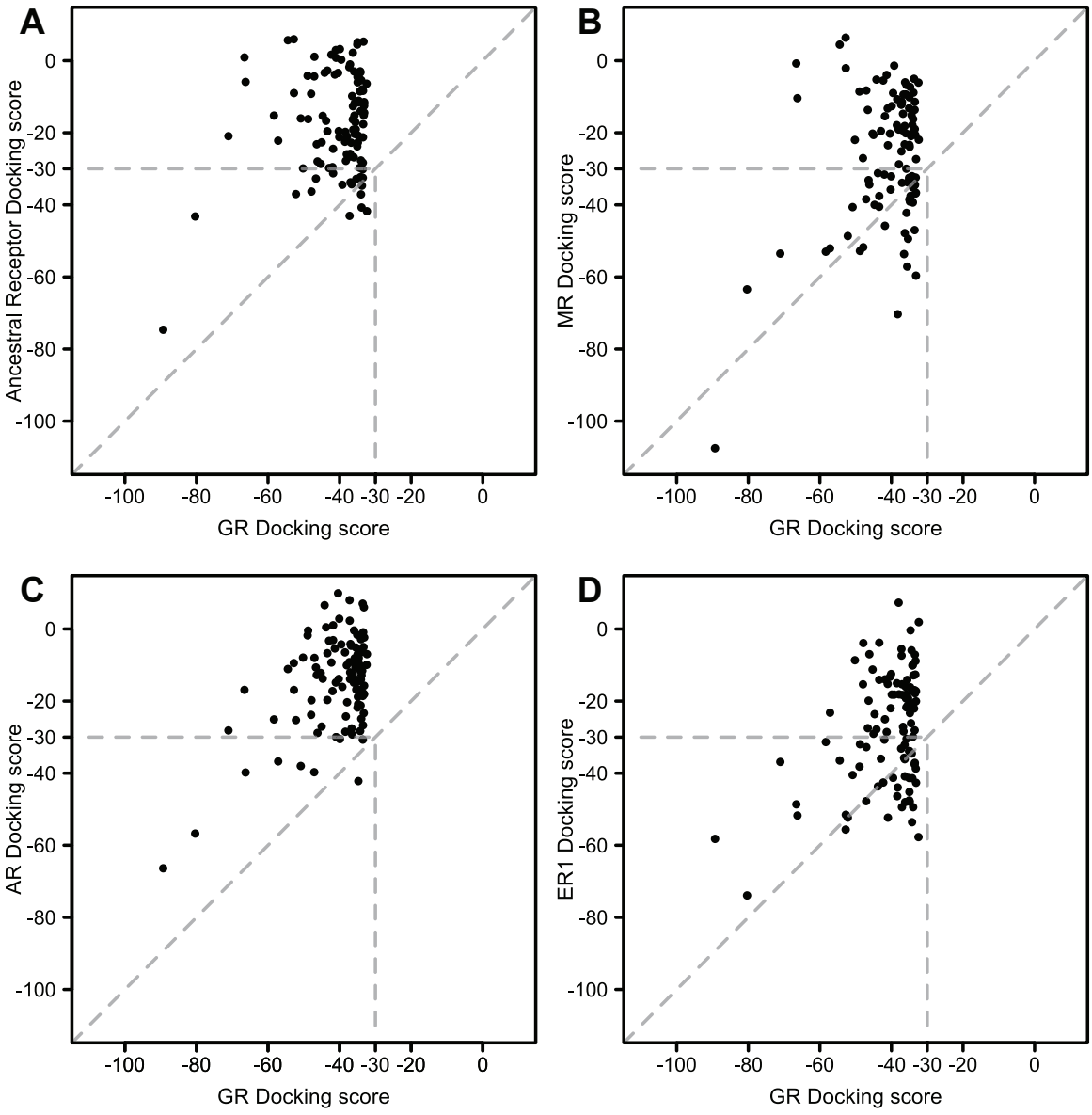


Figure 5.9 – Comparison of GR Docking Scores Versus Other SHR's
For parts **A) to D)**, the x-axes are all the GR docking scores and the y-axes are, respectively, the docking scores for AncR, MR, AR, and ER1. A dashed line along the diagonal indicates scores that are the same. Values above the diagonal line dock to GR better than the alternative SHR. A vertical dashed line indicates the weakest limit of GR docking and is replicated across the horizontal to indicate molecules that dock GR much better than the alternative SHR.

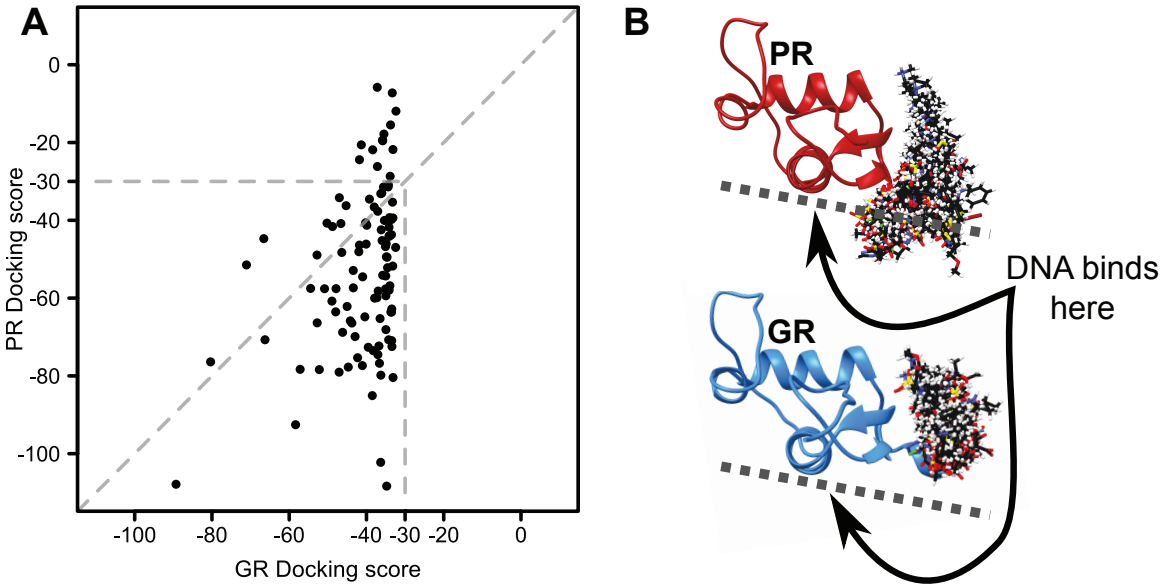


Figure 5.10 – Comparison of GR and PR Docking
Left: Comparison of PR and GR docking scores. The figure style is as in Figure 5.9. **Right:** The top image is of PR docked to the top hits of GR and the bottom image is of GR and its top hits. The dashed lines indicate roughly where DNA-binding occurs. Note, that the PR cluster of hits spreads into the area of DNA-binding.

the true positive hits. I have already ordered five of the lead compounds, and I have done ITC on one of the top hits, ZINC16188505 (negative result). I am retaining the small amounts of the other compounds so another graduate student, Emily Grasso, can detect any binding events by NMR. NMR is a better way to detect weak binding than ITC, and it also reveals the location of binding [81,322]. It is my hope that future decades of research will yield multiple DBD focused drugs.

5.5 Appendix

Table 5.2 – **List of the top 105 docked drug-like molecules**
The first column is the ZINC ID of a molecule (http://zinc.docking.org), the second column is the docking score (more negative is better), and the third column is the cluster that a molecule belongs to (see Figure 5.7). The hits are presented as best to worst scores.

ZINC-ID	Score	Cluster
ZINC94463428	-89.297371	1Aa
ZINC92451923	-80.351707	1B
ZINC92625407	-71.064026	1Aa
ZINC94221031	-66.585548	1B
ZINC94663954	-66.268997	1Aa
ZINC06002995	-58.345539	2A
ZINC91656096	-57.170429	1Aa
ZINC92181693	-54.463142	1B
ZINC94221032	-52.801006	1B
ZINC94221043	-52.762817	2Bb
ZINC92697378	-52.204086	1B
ZINC94875072	-50.855267	1B
ZINC77723060	-50.221661	2Bb
ZINC95413821	-48.941433	2Bb
ZINC94875222	-48.800243	1B
ZINC16188505	-47.96513	2Bb
ZINC20534671	-47.87841	1Ab
ZINC92697379	-47.085205	1B
ZINC86169897	-47.023155	1Ab
ZINC92625431	-46.618263	1Ab

Continued on next page

Table 5.2 – continued from previous page

ZINC-ID	Score	Cluster
ZINC92625447	-46.392277	1Ab
ZINC92625242	-46.17374	1Aa
ZINC77723236	-45.278801	2Ba
ZINC91294113	-45.016659	1Aa
ZINC32961786	-44.711979	1Ab
ZINC92862956	-44.188412	1Ab
ZINC94220831	-43.785179	1B
ZINC93091684	-43.423275	2Ba
ZINC92863429	-43.41222	1Ab
ZINC77723047	-42.949173	2Bb
ZINC92263294	-42.293922	2Bb
ZINC92626027	-41.980549	2Ab
ZINC10386625	-41.826099	1Ab
ZINC03291097	-41.816059	2Ba
ZINC77723291	-41.333954	2Ba
ZINC93909265	-41.059139	2Ba
ZINC72274798	-40.988071	1Ab
ZINC32962243	-40.422398	2Bb
ZINC95386276	-40.231777	1Aa
ZINC00177823	-40.089199	2Bb
ZINC91294020	-39.917656	1Ab

Continued on next page

Table 5.2 – continued from previous page

ZINC-ID	Score	Cluster
ZINC01760461	-39.537182	2Bb
ZINC92698201	-39.234394	1Aa
ZINC95372997	-38.490776	2Bb
ZINC92625948	-38.390018	1Ab
ZINC94663943	-38.246098	1B
ZINC78818954	-37.995689	2Bb
ZINC91753458	-37.897945	1B
ZINC91946165	-37.290211	2Ba
ZINC92451895	-37.211227	2A
ZINC92444548	-37.171913	1Aa
ZINC91006642	-37.081432	2Bb
ZINC92697469	-37.048656	1B
ZINC77723058	-36.960911	2Bb
ZINC94220832	-36.7738	1B
ZINC58330249	-36.640049	1Ab
ZINC94324685	-36.482742	1Aa
ZINC94220787	-36.319965	1B
ZINC93227479	-36.295589	1Aa
ZINC94221096	-36.270233	1B
ZINC94220830	-36.17984	1B
ZINC94463405	-36.02626	1Aa

Continued on next page

Table 5.2 – continued from previous page

ZINC-ID	Score	Cluster
ZINC01760558	-35.958752	2Ba
ZINC84584952	-35.852158	2A
ZINC71913809	-35.7957	2Bb
ZINC93091688	-35.793346	2Bb
ZINC92697466	-35.594639	2A
ZINC77723160	-35.48341	2Bb
ZINC92697428	-35.357677	2A
ZINC32961721	-35.07394	1Ab
ZINC67124154	-35.060715	2Bb
ZINC91293844	-35.007145	1Ab
ZINC65404186	-34.980572	1Ab
ZINC94875118	-34.969494	1B
ZINC37432814	-34.89727	1Ab
ZINC79755970	-34.839684	2Bb
ZINC92697520	-34.767067	1B
ZINC89484127	-34.669525	2Bb
ZINC91675334	-34.538677	1Ab
ZINC85606638	-34.390945	2Bb
ZINC94875196	-34.322971	1B
ZINC94221044	-34.251328	2Bb
ZINC21506288	-34.171871	1Ab

Continued on next page

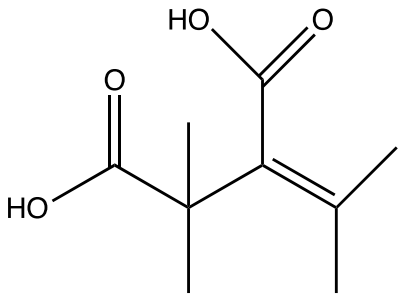
Table 5.2 – continued from previous page

ZINC-ID	Score	Cluster
ZINC94463404	-34.165104	1Aa
ZINC93091685	-34.131485	2Ba
ZINC94875195	-34.046158	1B
ZINC94555882	-34.018806	1Ab
ZINC25665460	-33.943085	1Ab
ZINC92625274	-33.863117	1Ab
ZINC06697830	-33.76355	2Bb
ZINC93909520	-33.679337	2Ba
ZINC37493642	-33.541897	2Bb
ZINC94071799	-33.524509	1Aa
ZINC93091691	-33.512066	2Ba
ZINC91675613	-33.473057	1Ab
ZINC32918007	-33.456585	2Ba
ZINC78756568	-33.409382	2Ba
ZINC86169941	-33.320339	2Bb
ZINC92978969	-33.253223	1Ab
ZINC89642392	-33.199356	2Ba
ZINC77723207	-33.143936	2Ba
ZINC77723015	-33.143299	2Bb
ZINC32917935	-33.123417	2Ba
ZINC89642327	-32.450695	2Ba

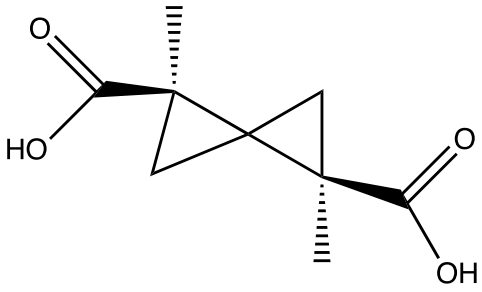
Continued on next page

Table 5.2 – continued from previous page

ZINC-ID	Score	Cluster
ZINC94236236	-32.368546	2Ba



ZINC 98090137; score = -121



ZINC 04728618; score = -48

Figure 5.11 – **Top two fragment screening hits**

Table 5.3 – **List of the top 61 docked fragment molecules**
The first column is the ZINC ID of a molecule (http://zinc.docking.org), the second column is the docking score (more negative is better). The hits are presented as best to worst scores.

ZINC-ID	Score
ZINC98090137	-121.158157
ZINC04728618	-48.097897
ZINC04244828	-38.479828
ZINC13540203	-37.980785
ZINC11920136	-35.53896
ZINC26443363	-35.399693
ZINC04029391	-34.076889

Continued on next page

Table 5.3 – continued from previous page

ZINC-ID	Score
ZINC16051320	-33.319141
ZINC01716647	-32.989086
ZINC08428163	-32.788197
ZINC13687237	-32.587322
ZINC04667621	-32.250809
ZINC03861122	-31.90427
ZINC13804228	-31.645563
ZINC38478095	-31.226097
ZINC04772668	-30.869179
ZINC00090750	-30.723333
ZINC20209252	-30.621716
ZINC21020932	-30.47389
ZINC98090213	-30.084116
ZINC02599010	-30.06218
ZINC19287259	-29.919485
ZINC85006743	-29.817389
ZINC00393354	-29.813595
ZINC82301007	-29.796078
ZINC05927438	-29.701788
ZINC12354630	-29.645439
ZINC02191273	-29.598709

Continued on next page

Table 5.3 – continued from previous page

ZINC-ID	Score
ZINC21948448	-29.519014
ZINC20123861	-29.340361
ZINC13408245	-29.337477
ZINC32500948	-29.260738
ZINC71976654	-29.013145
ZINC04811873	-28.87443
ZINC01532692	-28.866013
ZINC14000365	-28.80162
ZINC04473001	-28.799936
ZINC01698574	-28.572542
ZINC59724760	-28.530369
ZINC00393011	-28.500664
ZINC19735748	-28.452358
ZINC33380911	-28.418407
ZINC05319514	-28.359325
ZINC08549323	-28.282736
ZINC00895984	-28.130465
ZINC05178415	-28.121807
ZINC00134528	-27.921476
ZINC02507637	-27.881702
ZINC71781167	-27.789742

Continued on next page

Table 5.3 – continued from previous page

ZINC-ID	Score
ZINC02034090	-27.705805
ZINC00053391	-27.63501
ZINC34734485	-27.574903
ZINC04760065	-27.476723
ZINC19419105	-27.437019
ZINC00404362	-27.397343
ZINC71603262	-27.393894
ZINC12375595	-27.321114
ZINC00393695	-27.248365
ZINC19735652	-27.18272
ZINC01999514	-27.157654
ZINC00092914	-27.131645

The DOCK input file for grid docking: ligand_atom_file /home/jordan/Dock/dock6/GR_DBD/.JTW/-cluster1.mol2

limit_max_ligands no

skip_molecule no

read_mol_solvation no

calculate_rmsd no

use_database_filter no

orient_ligand yes

automated_matching yes

receptor_site_file /home/jordan/Dock/dock6/GR_DBD/DBD_cluster.sph

max_orientations 1000

critical_points no

chemical_matching no

use_ligand_spheres no

use_internal_energy yes

internal_energy_rep_exp 12

flexible_ligand yes

user_specified_anchor no

limit_max_anchors yes

max_anchor_num 2

min_anchor_size 5

num_anchors_orients_for_growth 2

pruning_use_clustering yes

pruning_max_orients 1000

pruning_clustering_cutoff 100

pruning_conformer_score_cutoff 100.0

use_clash_overlap no

write_growth_tree no

bump_filter no

score_molecules yes

contact_score_primary no

contact_score_secondary no

grid_score_primary yes

grid_score_secondary no

grid_score_rep_rad_scale 1

grid_score_vdw_scale 1

grid_score_es_scale 1

grid_score_grid_prefix /home/jordan/Dock/dock6/GR_DBD/.JTW/grid

multigrid_score_secondary no

dock3.5_score_secondary no

continuous_score_secondary no

descriptor_score_secondary no

gbsa_zou_score_secondary no

gbsa_hawkins_score_secondary no

SASA_descriptor_score_secondary no

amber_score_secondary no

minimize_ligand yes

minimize_anchor yes

minimize_flexible_growth yes

use_advanced_simplex_parameters no

simplex_max_cycles 1

simplex_score_converge 0.1

simplex_cycle_converge 1.0

simplex_trans_step 1.0

simplex_rot_step 0.1

simplex_tors_step 10.0

simplex_anchor_max_iterations 500

simplex_grow_max_iterations 500

simplex_grow_tors_premin_iterations 0

simplex_random_seed 0

simplex_restraint_min no

atom_model all

vdw_defn_file /home/jordan/Dock/dock6/parameters/vdw_AMBER_parm99.defn

flex_defn_file /home/jordan/Dock/dock6/parameters/flex.defn

flex_drive_file /home/jordan/Dock/dock6/parameters/flex_drive.tbl

ligand_outfile_prefix Fragments_anchor

write_orientations no

num_scored_conformers 1

rank_ligands yes

max_ranked_ligands 400

******

The DOCK input file for AMBER docking:

ligand_atom_file AMBER_ligand_test.amber_score.mol2

limit_max_ligands no

skip_molecule no

read_mol_solvation no

calculate_rmsd no

use_database_filter no

orient_ligand no

use_internal_energy no

flexible_ligand no

bump_filter no

score_molecules yes

contact_score_primary no

contact_score_secondary no

grid_score_primary no

grid_score_secondary no

multigrid_score_primary no

multigrid_score_secondary no

dock3.5_score_primary no

dock3.5_score_secondary no

continuous_score_primary no

continuous_score_secondary no

descriptor_score_primary no

descriptor_score_secondary no

gbsa_zou_score_primary no

gbsa_zou_score_secondary no

gbsa_hawkins_score_primary no

gbsa_hawkins_score_secondary no

SASA_descriptor_score_primary no

SASA_descriptor_score_secondary no

amber_score_primary yes

amber_score_secondary no

amber_score_receptor_file_prefix DBD

amber_score_movable_region everything

amber_score_minimization_rmsgrad 0.01

amber_score_before_md_minimization_cycles 100

amber_score_md_steps 3000

amber_score_after_md_minimization_cycles 100

amber_score_gb_model 5

amber_score_nonbonded_cutoff 18.0

amber_score_temperature 300.0

amber_score_abort_on_unprepped_ligand yes

ligand_outfile_prefix AMBER_output_everything

write_orientations no

num_scored_conformers 1

rank_ligands yes

max_ranked_ligands 100

Bibliography

[1] S. Adachi, M. Okuno, R. Matsushima-Nishiwaki, Y. Takano, S. Kojima, S. L. Friedman, H. Moriwaki, and Y. Okano. Phosphorylation of retinoid x receptor suppresses its ubiquitination in human hepatocellular carcinoma. Hepatology, 35(2):332–40, Feb 2002.

[2] G. Adair, A. Bock, and H. jr Field. The hemoglobin system: Vi. the oxygen dissociation curve of hemoglobin. Journal of Biological Chemistry, 63:529–545, 1925.

[3] M. A. Y. Adell, G. F. Vogel, M. Pakdel, M. Müller, H. Lindner, M. W. Hess, and D. Teis. Coordinated binding of vps4 to escrt-iii drives membrane neck constriction during mvb vesicle formation. J Cell Biol, 2014.

[4] R. Adhikary, Y. X. Tan, J. Liu, J. Zimmermann, M. Holcomb, C. Yvellez, P. E. Dawson, and F. E. Romesberg. Conformational heterogeneity and dna recognition by the morphogen bicoid. Biochemistry, 2017.

[5] M. Agromayor, N. Soler, A. Caballe, T. Kueck, S. M. Freund, M. D. Allen, M. Bycroft, O. Perisic, Y. Ye, B. McDonald, H. Scheel, K. Hofmann, S. J. Neil, J. Martin-Serrano, and R. L. Williams. The ubap1 subunit of escrt-i interacts with ubiquitin via a souba domain. Structure, 2012.

[6] S. L. Alam, C. Langelier, F. G. Whitby, S. Koirala, H. Robinson, C. P. Hill, and W. I. Sundquist. Structural basis for ubiquitin recognition by the human escrt-ii eap45 glue domain. Nat Struct Mol Biol, 2006.

[7] K. Alexandrov, H. Horiuchi, O. Steele-Mortimer, M. C. Seabra, and M. Zerial. Rab escort protein-1 is a multifunctional protein that accompanies newly prenylated rab proteins to their target membranes. EMBO J, 1994.

[8] N. Ali, L. Zhang, S. Taylor, A. Mironov, S. Urbe, and P. Woodman. Recruitment of ubpy and escrt exchange drive hd-ptp-dependent sorting of egfr to the mvb. Curr Biol, 2013.

[9] G. F. Allan, X. Leng, S. Y. Tsai, N. L. Weigel, D. P. Edwards, M. J. Tsai, and B. W. O'Malley. Hormone and antihormone induce distinct conformational changes which are central to steroid receptor activation. J Biol Chem, 267(27):19513–20, Sep 25 1992.

[10] W. Allen, T. Balius, S. Mukherjee, S. Brozell, D. Moustakas, P. Lang, D. Case, I. Kuntz, and R. Rizzo. Dock 6: Impact of new features and current ddocki performance. J Comput Chem, 2015.

[11] T. Almlof, A. E. Wallberg, J.-A. Gustafsson, and A. P. H. Wright. Role of important hydrophobic amino acids in the interaction between the glucocorticoid receptor τ1-core activation domain and target factors. Biochemistry, 1998.

[12] I. Amit, L. Yakir, M. Katz, Y. Zwang, M. D. Marmor, A. Citri, K. Shtiegman, I. Alroy, S. Tuvia, Y. Reiss, E. Roubini, M. Cohen, R. Wides, E. Bacharach, U. Schubert, and Y. Yarden. Tal, a tsg101-specific e3 ubiquitin ligase, regulates receptor endocytosis and retrovirus budding. Genes Dev, 18(14):1737–52, Jul 15 2004.

[13] A. Amsterdam, K. Tajima, and R. Sasson. Cell-specific regulation of apoptosis by glucocorticoids: implication to their anti-inflammatory action. Biochem Pharmacol, 64(5-6):843–50, Sep 2002.

[14] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. H++ 3.0: automating pk prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res, 40(Web Server issue):W537–41, Jul 2012.

[15] B. Apostolovic and H.-A. Klok. pH-sensitivity of the E3/K3 heterodimeric coiled coil. Biomacromolecules, 9(11):3173–3180, 2008.

[16] A. Aranda and A. Pascual. Nuclear hormone receptors and gene expression. Physiol Rev, 81(3):1269–304, Jul 2001.

[17] G. Arents, R. W. Burlingame, B.-C. Wang, W. E. Love, and E. N. Moudrianakis. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proceedings of the National Academy of Sciences, 88(22):10148–10152, 1991.

[18] G. Armony, E. Jacob, T. Moran, Y. Levin, T. Mehlmand, Y. Levy, and D. Fass. Cross-linking reveals laminin coiled-coil architecture. PNAS, 2016.

[19] V. K. Arora, E. Schenkein, R. Murali, S. K. Subudhi, J. Wongvipat, M. D. Balbas, N. Shah, L. Cai, E. Efstathiou, C. Logothetis, D. Zheng, and C. L. Sawyers. Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. Cell, 155(6):1309–22, Dec 5 2013.

[20] I. F. Azmi, B. A. Davies, J. Xiao, M. Babst, Z. Xu, and D. J. Katzmann. Escrt-iii family members stimulate vps4 atpase activity directly or via vta1. Developmental Cell, 2008.

[21] M. Babst, D. J. Katzmann, E. J. Estepa-Sabal, T. Meerloo, and S. D. Emr. Escrt-iii: An endosome-associated heterooligomeric protein complex required for mvb sorting. Developmental Cell, 2002.

[22] M. Babst, D. J. Katzmann, W. B. Snyder, B. Wendland, and S. D. Emr. Endosome-associated complex, escrt-ii, recruits transport machinery for protein sorting at the multivesicular body. Developmental Cell, 2002.

[23] M. Babst, G. Odorizzi, E. J. Estepa, and S. D. Emr. Mammalian tumor susceptibility gene 101 (tsg101) and the yeast homologue, vps23p, both function in late endosomal trafficking. Traffic, 1(3):248–58, Mar 2000.

[24] M. Babst, T. K. Sato, L. M. Banta, and S. D. Emr. Endosomal transport function in yeast requires a novel aaa-type atpase, vps4p. EMBO J, 1997.

[25] M. Babst, B. Wendland, E. J. Estepa, and S. D. Emr. The vps4p aaa atpase regulates membrane association of a vps protein complex required for normal endosome function. EMBO J, 1998.

[26] M. K. Bagchi, J. F. Elliston, S. Y. Tsai, D. P. Edwards, M. J. Tsai, and B. W. O'Malley. Steroid hormone-dependent interaction of human progesterone receptor with its target enhancer element. Mol Endocrinol, 2(12):1221–9, Dec 1988.

[27] M. F. Baietti, Z. Zhang, E. Mortier, A. Melchior, G. Degeest, A. Geeraerts, Y. Ivarsson, F. Depoortere, C. Coomans, E. Vermeiren, P. Zimmermann, and G. David. Syndecan–syntenin–alix regulates the biogenesis of exosomes. Nat Cell Biol, 2012.

[28] D. L. Bain, M. A. Franden, J. L. McManaman, G. S. Takimoto, and K. B. Horwitz. The n-terminal region of the human progesterone a-receptor. structural analysis and the influence of the dna binding domain. J Biol Chem, 275(10):7313–20, Mar 10 2000.

[29] D. L. Bain, M. A. Franden, J. L. McManaman, G. S. Takimoto, and K. B. Horwitz. The n-terminal region of human progesterone b-receptors: biophysical and biochemical comparison to a-receptors. J Biol Chem, 276(26):23825–31, Jun 29 2001.

[30] D. L. Bain, Q. Yang, K. D. Connaghan, J. P. Robblee, M. T. Miura, G. D. Degala, J. R. Lambert, and N. K. Maluf. Glucocorticoid receptor-dna interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity. J Mol Biol, 422(1):18–32, Sep 7 2012.

[31] K. Bajaj, G. Chakshusmathi, K. Bachhawat-Sikder, A. Surolia, and R. Varadarajan. Thermodynamic characterization of monomeric and dimeric forms of ccdb (controller of cell division or death b protein). Biochem J, 2004.

[32] M. Bajorek, H. L. Schubert, J. McCullough, C. Langelier, D. M. Eckert, W.-M. B. Stubblefield, N. T. Uter, D. G. Myszka, C. P. Hill, and W. I. Sundquist. Structural basis for escrt-iii protein autoinhibition. Nature Structural & Molecular Biology, 2009.

[33] R. L. Baldwin. Temperature dependence of the hydrophobic interaction in protein folding. Proceedings of the National Academy of Sciences, 83(21):8069–8072, 1986.

[34] T. Balla. Phosphoinositides: Tiny lipids with giant impact on cell regulation. Physiol Rev, 2013.

[35] D. W. Banner, M. Kokkinidis, and D. Tsernoglou. Structure of the ColE1 rop protein at 1.7 \AA resolution. Journal of molecular biology, 196(3):657–675, 1987.

[36] J. Barcroft and A. V. Hill. The nature of oxyhaemoglobin, with a note on its molecular weight. J Physiol, 39(6):411–28, Mar 8 1910.

[37] M. A. Barry and A. Eastman. Endonuclease activation during apoptosis: the role of cytosolic Ca2+ and pH. Biochemical and biophysical research communications, 186(2):782–789, 1992.

[38] I. V. Baskakov, R. Kumar, G. Srinivasan, Y. S. Ji, D. W. Bolen, and E. B. Thompson. Trimethylamine n-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. The Journal of biological chemistry, 274(16):10693–6, Apr 16 1999.

[39] P. B. Becker, B. Gloss, W. Schmid, U. Strahle, and G. Schutz. In vivo protein-dna interactions in a glucocorticoid response element require the presence of the hormone. Nature, 324(6098):686–8, Dec 18-31 1986.

[40] J. M. Beekman, G. F. Allan, S. Y. Tsai, M. J. Tsai, and B. W. O'Malley. Transcriptional activation by the estrogen receptor requires a conformational change in the ligand binding domain. Mol Endocrinol, 7(10):1266–74, Oct 1993.

[41] I. K. Bender, Y. Cao, and N. Z. Lu. Determinants of the heightened activity of glucocorticoid receptor translational isoforms. Endocrinology, 2013.

[42] Y. Benjamini, Y Hochberg. Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological, 1995.

[43] H. Berglund, M. Wolf-Watz, T. Lundback, S. van den Berg, and T. Hard. Structure and dynamics of the glucocorticoid receptor dna-binding domain: comparison of wild type and a mutant with altered specificity. Biochemistry, 36(37):11188–97, Sep 16 1997.

[44] P. Birth, S. Schone, U. Stelzl, and S. H. Meijsing. Identification and characterization of batf3 as a context-specific coactivator of the glucocorticoid receptor. PLOS ONE, 2017.

[45] C. Bissig, M. Lenoir, M.-C. Velluz, I. Kufareva, R. Abagyan, M. Overduin, and J. Gruenberg. Viral infection controlled by a calcium-dependent lipid-binding module in alix. Developmental Cell, 2013.

[46] J. Blackford, J. A., C. Guo, R. Zhu, E. J. Dougherty, C. C. Chow, and J. Simons, S. S. Identification of location and kinetically defined mechanism of cofactors and reporter genes in the cascade of steroid-regulated transactivation. J Biol Chem, 287(49):40982–95, Nov 30 2012.

[47] E. V. Bocharov, P. E. Bragin, K. V. Pavlov, O. V. Bocharova, K. S. Mineev, A. A. Polyansky, P. E. Volynsky, R. G. Efremov, and A. S. Arseniev. The conformation of the epidermal growth factor receptor transmembrane domain dimer dynamically adapts to the local membrane environment. Biochemistry, 2017.

[48] D. W. Bolen and I. V. Baskakov. The osmophobic effect: natural selection of a thermodynamic force in protein folding. J Mol Biol, 310(5):955–63, Jul 27 2001.

[49] A. M. Bonvin, H. Vis, J. N. Breg, M. J. Burgering, R. Boelens, and R. Kaptein. Nuclear magnetic resonance solution structure of the Arc repressor using relaxation matrix calculations. Journal of molecular biology, 236(1):328–341, 1994.

[50] O. Boudker, M. J. Todd, and E. Freire. The structural stability of the co-chaperonin GroES. Journal of molecular biology, 272(5):770–779, 1997.

[51] E. Boura and J. H. Hurley. Structural basis for membrane targeting by the mvb12-associated β-prism domain of the human escrt-i mvb12 subunit. PNAS, 2012.

[52] E. Boura, B. Rozycki, H. S. Chung, D. Z. Herrick, B. Canagarajah, D. S. Cafiso, W. A. Eaton, G. Hummer, and J. H. Hurley. Solution structure of the escrt-i and

-ii supercomplex: Implications for membrane budding and scission. Structure, 2012.

[53] W. Bourguet, M. Ruff, P. Chambon, H. Gronemeyer, and D. Moras. Crystal structure of the ligand-binding domain of the human nuclear receptor rxr-alpha. Nature, 375(6530):377–82, Jun 1 1995.

[54] K. Bowers, J. Lottridge, S. B. Helliwell, L. M. Goldthwaite, J. P. Luzio, and T. H. Stevens. Protein–protein interactions of escrt complexes in the yeast saccharomyces cerevisiae. Traffic, 2004.

[55] J. U. Bowie and R. T. Sauer. Equilibrium dissociation and unfolding of the Arc repressor dimer. Biochemistry, 28(18):7139–7143, 1989.

[56] S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. DiMaio, J. H. Pereira, and others. De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity. Science, 352(6286):680–687, 2016.

[57] S. Breinig, J. Kervinen, L. Stith, A. S. Wasson, R. Fairman, A. Wlodawer, A. Zdanov, and E. K. Jaffe. Control of tetrapyrrole biosynthesis by alternate quaternary forms of porphobilinogen synthase. Nature Structural & Molecular Biology, 10(9):757–763, 2003.

[58] J. Brodie and I. J. McEwan. Intra-domain communication between the n-

terminal and dna-binding domains of the androgen receptor: modulation of androgen response element dna binding. J Mol Endocrinol, 34(3):603–15, Jun 2005.

[59] A. M. Brzozowski, A. C. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson, and M. Carlquist. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature, 389(6652):753–8, Oct 16 1997.

[60] N. J. Buchkovich, W. M. Henne, S. Tang, and S. D. Emr. Essential n-terminal insertion motif anchors the escrt-iii filament during mvb vesicle formation. Developmental Cell, 2013.

[61] M. Buljan, G. Chalancon, S. Eustermann, G. P. Wagner, M. Fuxreiter, A. Bateman, and M. M. Babu. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell, 46(6):871–83, Jun 29 2012.

[62] T. D. Bunney, P. A. C. Watkins, A. F. Beven, P. J. Shaw, L. E. Hernandez, G. P. Lomonossoff, M. Shanks, J. Peart, and B. K. Drobak. Association of phosphatidylinositol 3-kinase with nuclear transcription sites in higher plants. The Plant Cell, 2000.

[63] S. Burgdorf, P. Leister, and K. H. Scheidtmann. Tsg101 interacts with apoptosis-antagonizing transcription factor and enhances androgen receptor-

mediated transcription by promoting its monoubiquitination. J Biol Chem, 2004.

[64] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage dna. Science, 1994.

[65] A. Caballe, D. M. Wenzel, M. Agromayor, S. L. Alam, J. J. Skalicky, M. Kloc, J. G. Carlton, L. Labrador, W. I. Sundquist, and J. Martin-Serrano. Ulk3 regulates cytokinetic abscission by phosphorylating escrt-iii proteins. eLife, 2015.

[66] K. M. Campbell, A. J. Sholders, and K. J. Lumb. Contribution of buried lysine residues to the oligomerization specificity and stability of the Fos coiled coil. Biochemistry, 41(15):4866–4871, 2002.

[67] F. Campelo, H. T. McMahon, and M. M. Kozlov. The hydrophobic insertion mechanism of membrane curvature generation by proteins. Biophysical journal, 2008.

[68] G. Candiano, M. Bruschi, L. Musante, L. Santucci, G. M. Ghiggeri, B. Carnemolla, P. Orecchia, L. Zardi, and P. G. Righetti. Blue silver: a very sensitive colloidal coomassie g-250 staining for proteome analysis. Electrophoresis, 25(9):1327–33, May 2004.

[69] Y. Cao, I. K. Bender, A. K. Konstantinidis, S. C. Shin, C. M. Jewell, J. A.

Cidlowski, R. P. Schleimer, and N. Z. Lu. Glucocorticoid receptor translational isoforms underlie maturational stage-specific glucocorticoid sensitivities of dendritic cells in mice and humans. Blood, 121(9):1553–62, Feb 28 2013.

[70] Y. Cao, A. Charisi, L. C. Cheng, T. Jiang, and T. Girke. Chemminer: a compound mining framework for r. Bioinformatics, 24(15):1733–1734, Aug 1 2008.

[71] A. G. Cashikar, S. Shim, R. Roth, M. R. Maldazys, J. E. Heuser, and P. I. Hanson. Structure of cellular escrt-iii spirals and their relationship to hiv budding. eLife, 2014.

[72] T. Castro-Gomes, A. Koushik, and N. Andrews. Escrt: Nipping the wound in the bud? Trends in Biochemical Sciences, 2014.

[73] L. Christ, E. M. Wenzel, K. Liestol, C. Raiborg, C. Campsteijn, and H. Stenmark. Alix and escrt-i/ii function as parallel escrt-iii recruiters in cytokinetic abscission. J Cell Biol, 2016.

[74] H.-H. Chua, H.-H. Lee, S.-S. Chang, C.-C. Lu, T.-H. Yeh, T.-Y. Hsu, T.-H. Cheng, J.-T. Cheng, M.-R. Chen, and C.-H. Tsai. Role of the tsg101 gene in epstein-barr virus late gene transcription. Journal of Virology, 2007.

[75] A. J. Copik, M. S. Webb, A. L. Miller, Y. Wang, R. Kumar, and E. B. Thompson.

Activation function 1 of glucocorticoid receptor binds tata-binding protein in vitro and in vivo. Mol Endocrinol, 20(6):1218–30, Jun 2006.

[76] L. Corless, C. M. Crump, S. D. C. Griffin, and M. Harris. Vps4 and the escrt-iii complex are required for the release of infectious hepatitis c virus particles. Journal of General Virology, 2010.

[77] P. A. Costeas and J. M. Chinsky. Glucocorticoid regulation of branched-chain α-ketoacid dehydrogenase e2 subunit gene expression. Biochem J, 2000.

[78] M. K. Crowder, C. D. Seacrist, and R. D. Blind. Phospholipid regulation of the nuclear receptor superfamily. Advances in Biological Regulation, 2017.

[79] K. Dahlman-Wright, H. Baumann, I. J. McEwan, T. Almlof, A. P. Wright, J. A. Gustafsson, and T. Hard. Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. Proc Natl Acad Sci U S A, 92(5):1699–703, Feb 28 1995.

[80] M. D. Daily and J. J. Gray. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. PLoS Comput Biol, 5(2):e1000293, 2009.

[81] C. Dalvit, A. D. Gossert, J. Coutant, and M. Piotto. Rapid acquisition of 1h and 19f nmr experiments for direct and competition ligand-based screening. Magn Reson Chem, 49(4):199–202, Apr 2011.

[82] J. A. D'Aquino, J. Gómez, V. J. Hilser, K. H. Lee, L. M. Amzel, and E. Freire. The magnitude of the backbone conformational entropy change in protein folding. Proteins: Structure, Function, and Bioinformatics, 25(2):143–156, 1996.

[83] B. A. Davies, I. F. Azmi, J. Payne, A. Shestakova, B. F. Horazdovsky, M. Babst, and D. J. Katzmann. Coordination of substrate binding and atp hydrolysis in vps4-mediated escrt-iii disassembly. Mol Biol Cell, 2010.

[84] C. L. Day and T. Alber. Crystal structure of the amino-terminal coiled-coil domain of the APC tumor suppressor. Journal of molecular biology, 301(1):147–156, 2000.

[85] D. Dayde, M. Guerard, P. Perron, A.-S. Hatat, C. Barrial, B. Eymin, and S. Gazzeri. Nuclear trafficking of egfr by vps34 represses arf expression to promote lung tumor cell survival. Oncogene, 2016.

[86] P. de Lange, C. M. Segeren, J. W. Koper, E. Wiemer, P. Sonneveld, A. O. Brinkmann, A. White, I. J. Brogan, F. H. de Jong, and S. W. Lamberts. Expression in hematological malignancies of a glucocorticoid receptor splice variant that augments glucocorticoid receptor-mediated effects in transfected cells. Cancer Res, 61(10):3937–41, May 15 2001.

[87] C. Dekker, B. de Kruijff, and P. Gros. Crystal structure of SecB from Escherichia coli. Journal of structural biology, 144(3):313–319, 2003.

[88] F. E. Dela Cruz, D. R. Kirsch, and J. N. Heinrich. Transcriptional activity of drosophila melanogaster ecdysone receptor isoforms and ultraspiracle in saccharomyces cerevisiae. J Mol Endocrinol, 24(2):183–91, Apr 2000.

[89] C. M. Denais, R. M. Gilbert, P. Isermann, A. L. McGregor, M. te Lindert, B. Weigelin, P. M. Davidson, P. Friedl, K. Wolf, and J. Lammerding. Nuclear envelope rupture and repair during cancer cell migration. Science, 2016.

[90] E. Diamanti-Kandarakis, J. P. Bourguignon, L. C. Giudice, R. Hauser, G. S. Prins, A. M. Soto, R. T. Zoeller, and A. C. Gore. Endocrine-disrupting chemicals: an endocrine society scientific statement. Endocrine reviews, 30(4):293–342, Jun 2009.

[91] R. E. Dickerson. X-ray studies of protein mechanisms. Annu Rev Biochem, 1972.

[92] M. J. Dobro, R. Y. Samson, Z. Yu, J. McCullough, H. J. Ding, P. L.-G. Chong, S. D. Bell, and G. J. Jensen. Electron cryotomography of escrt assemblies and dividing sulfolobus cells suggests that spiraling filaments are involved in membrane scission. Mol Biol Cell, 2013.

[93] M. R. Dores, N. J. Grimsey, F. Mendez, and J. Trejo. Alix regulates the ubiquitin-independent lysosomal sorting of the p2y1 purinergic receptor via a ypx3l motif. PLoS ONE, 2016.

[94] J. J. Dougherty, R. K. Puri, and D. O. Toft. Phosphorylation in vivo of chicken oviduct progesterone receptor. J Biol Chem, 257(23):14226–30, Dec 10 1982.

[95] D. P. Dowlatshahi, V. Sandrin, S. Vivona, T. A. Shaler, S. E. Kaiser, F. Melandri, W. I. Sundquist, and R. R. Kopito. Alix is a lys63-specific polyubiquitin binding protein that functions in retrovirus budding. Developmental Cell, 2012.

[96] A. Doyotte, A. Mironov, E. McKenzie, and P. Woodman. The bro1-related protein hd-ptp/ptpn23 is required for endosomal cargo sorting and multivesicular body morphogenesis. PNAS, 2008.

[97] A. Dragan, S. Potekhin, A. Sivolob, M. Lu, and P. L. Privalov. Kinetics and thermodynamics of the unfolding and refolding of the three-stranded alpha-helical coiled coil, lpp-56. Biochemistry, 2004.

[98] O. K. Dudko, G. Hummer, and A. Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. PNAS, 2008.

[99] J. Dunbrack, R. L. Rotamer libraries in the 21st century. Curr Opin Struct Biol, 12(4):431–40, Aug 2002.

[100] H. Edelhoch. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry, 6(7):1948–54, Jul 1967.

[101] R. Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res, 2004.

[102] N. Elia, R. Sougrat, T. A. Spurlin, J. H. Hurley, and J. Lippincott-Schwartz. Dynamics of endosomal sorting complex required for transport (escrt) machinery during cytokinesis and its role in abscission. Proceedings of the National Academy of Sciences of the United States of America, 108(12):4846–4851, Mar 22 2011.

[103] R. Fairman, R. Beran-Steed, S. Anthony-Cahill, J. Lear, W. r. Stafford, W. Degrado, P. Benfield, and S. Brenner. Multiple oligomeric states regulate the dna binding of helix-loop-helix peptides. PNAS, 1993.

[104] S. Felder, K. Miller, G. Moehren, A. Ullrich, J. Schlessinger, and C. R. Hopkins. Kinas activity controls the sorting of the epidermal growth factor receptor within the multivesicular body. Cell, 1990.

[105] K. M. Ferguson, M. B. Berger, J. M. Mendrola, H.-S. Cho, D. J. Leahy, and M. A. Lemmon. Egf activates its receptor by removing interactions that autoinhibit ectodomain dimerization. Mol Cell, 2003.

[106] A. C. M. Ferreon, J. C. Ferreon, P. E. Wright, and A. A. Deniz. Modulation of allostery by protein intrinsic disorder. Nature, 498(7454):390–394, 2013.

[107] M. Filimonenko, S. Stuffers, C. Raiborg, A. Yamamoto, L. Malerod, E. M. Fisher, A. Isaacs, A. Brech, H. Stenmark, and A. Simonsen. Functional multivesicular bodies are required for autophagic clearance of protein aggregates associated with neurodegenerative disease. J Cell Biol, 2007.

[108] K. Fischer, S. M. Kelly, K. Watt, N. C. Price, and I. J. McEwan. Conformation of the mineralocorticoid receptor n-terminal domain: evidence for induced and stable structure. Mol Endocrinol, 24(10):1935–48, Oct 2010.

[109] B. M. Forman, J. Chen, and R. M. Evans. Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. Proc Natl Acad Sci U S A, 94(9):4312–7, Apr 29 1997.

[110] E. Freire. Statistical thermodynamic analysis of the heat capacity function associated with protein folding-unfolding transitions. Comments Molecular Cellular Biophysics, 6:123–140, 1989.

[111] E. Freire and R. L. Biltonen. Statistical mechanical deconvolution of thermal transitions in macromolecules. I. Theory and application to homogeneous systems. Biopolymers, 17(2):463–479, 1978.

[112] M. Fu, C. Wang, A. T. Reutens, J. Wang, R. H. Angeletti, L. Siconolfi-Baez, V. Ogryzko, M. L. Avantaggiati, and R. G. Pestell. p300 and p300/camp-response element-binding protein-associated factor acetylate the androgen receptor at sites governing hormone-dependent transactivation. J Biol Chem, 275(27):20853–60, Jul 7 2000.

[113] I. Fyfe, A. L. Schuh, J. M. Edwardson, and A. Audhya. Association of the endosomal sorting complex escrt-ii with the vps20 subunit of escrt-iii generates

a curvature-sensitive complex capable of nucleating escrt-iii filaments. J Biol Chem, 2011.

[114] C. Gancedo and C.-L. Flores. Moonlighting proteins in yeast. Microbiology and Molecular Biology Reviews, 2008.

[115] C. Gaoa, X. Zhuanga, Y. Cuia, X. Fua, Y. Hea, Q. Zhaoa, Y. Zenga, J. Shena, M. Luoa, and L. Jiang. Dual roles of an arabidopsis escrt component free1 in regulating vacuolar protein transport and autophagic degradation. PNAS, 2015.

[116] V. Garcia-Carpizo, L. Ruiz-Llorente, M. Fraga, and A. Aranda. The growing role of gene methylation on endocrine function. J Mol Endocrinol, 47(2):R75–89, Oct 2011.

[117] J. E. Garrus, U. K. von Schwedler, O. W. Pornillos, S. G. Morham, K. H. Zavitz, H. E. Wang, D. A. Wettstein, K. M. Stray, M. Cote, R. L. Rich, D. G. Myszka, and W. I. Sundquist. Tsg101 and the vacuolar protein sorting pathway are essential for hiv-1 budding. Cell, 2001.

[118] H. Garside, A. Stevens, S. Farrow, C. Normand, B. Houle, A. Berry, B. Maschera, and D. Ray. Glucocorticoid ligands specify different interactions with nf-kappab by allosteric effects on the glucocorticoid receptor dna binding domain. J Biol Chem, 279(48):50050–9, Nov 26 2004.

[119] A. C. Gee, K. E. Carlson, P. G. Martini, B. S. Katzenellenbogen, and J. A. Katzenellenbogen. Coactivator peptides have a differential stabilizing effect on the binding of estrogens and antiestrogens with the estrogen receptor. Mol Endocrinol, 13(11):1912–23, Nov 1999.

[120] K. A. Gelato, M. Tauber, M. S. Ong, S. Winter, K. Hiragami-Hamada, J. Sindlinger, A. Lemak, Y. Bultsma, S. Houliston, D. Schwarzer, N. Divecha, C. H. Arrowsmith, and W. Fischle. Accessibility of different histone h3-binding domains of uhrf1 is allosterically regulated by phosphatidylinositol 5-phosphate. Mol Cell, 2014.

[121] A. N. Gerber, K. Masuno, and M. I. Diamond. Discovery of selective glucocorticoid receptor modulators by multiplexed reporter screening. Proc Natl Acad Sci U S A, 106(12):4929–34, Mar 24 2009.

[122] C. Geserick, H. A. Meyer, K. Barbulescu, and B. Haendler. Differential modulation of androgen receptor action by deoxyribonucleic acid response elements. Mol Endocrinol, 17(9):1738–50, Sep 2003.

[123] S. Ghazi-Tabatabai, S. Saksena, J. M. Short, A. V. Pobbati, D. B. Veprintsev, R. A. Crowther, S. D. Emr, E. H. Egelman, and R. L. Williams. Structure and disassembly of filaments formed by the escrt-iii subunit vps24. Structure, 2008.

[124] D. J. Gill, H. Teo, J. Sun, O. Perisic, D. B. Veprintsev, S. D. Emr, and R. L.

Williams. Structural insight into the ESCRT-I/-II link and its role in MVB trafficking. The EMBO journal, 26(2):600–612, 2007.

[125] D. J. Gillooly, I. C. Morrow, M. Lindsey, R. Gould, N. J. Bryant, J.-M. Gaullier, R. G. Parton, and H. Stenmark. Localization of posphatidylinositol 3-phosphate in yeast and mammalian cells. EMBO J, 2000.

[126] C. K. Glass, J. M. Holloway, O. V. Devary, and M. G. Rosenfeld. The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone and estrogen response elements. Cell, 54(3):313–23, Jul 29 1988.

[127] R. Goila-Gaur, D. G. Demirov, J. M. Orenstein, A. Ono, and E. O. Freed. Defects in human immunodeficiency virus budding and endosomal sorting induced by tsg101 overexpression. J Virol, 2003.

[128] J. Gomez, V. J. Hilser, D. Xie, and E. Freire. The heat capacity of proteins. Proteins: Structure, Function, and Bioinformatics, 22(4):404–412, 1995.

[129] H. Gouda, H. Torigoe, A. Saito, M. Sato, Y. Arata, and I. Shimada. 3-dimensional solution structure of the b-domain of staphylococcal protein-a - comparisons of the solution and crystal-structures. Biochemistry, 1992.

[130] N. S. Green, E. Reisler, and K. N. Houk. Quantitative evaluation of the lengths

of homobifunctional protein cross-linking reagents used as molecular rulers. Protein Sci, 2001.

[131] N. Greenfield, V. Vijayanathan, T. J. Thomas, M. A. Gallo, and T. Thomas. Increase in the stability and helical content of estrogen receptor alpha in the presence of the estrogen response element: analysis by circular dichroism spectroscopy. Biochemistry, 40(22):6646–52, Jun 5 2001.

[132] N. J. Greenfield and G. D. Fasman. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry, 1969.

[133] J. Guizetti, L. Schermelleh, J. Mäntler, S. Maar, I. Poser, H. Leonhardt, T. Müller-Reichert, and D. W. Gerlich. Abscission involves helices of escrt-iii–dependent filaments. Sience, 2011.

[134] A. Hagendorf, J. W. Koper, F. H. de Jong, A. O. Brinkmann, S. W. Lamberts, and R. A. Feelders. Expression of the human glucocorticoid receptor splice variants alpha, beta, and p in peripheral blood mononuclear leukocytes in healthy controls and in patients with hyper- and hypocortisolism. J Clin Endocrinol Metab, 90(11):6237–43, Nov 2005.

[135] Y. Hagihara, M. Oobatake, and Y. Goto. Thermal unfolding of tetrameric melittin: comparison with the molten globule state of cytochrome c. Protein Science, 3(9):1418–1429, 1994.

[136] S. B. Hake, B. A. Garcia, E. M. Duncan, M. Kauer, G. Dellaire, J. Shabanowitz, D. P. Bazett-Jones, C. D. Allis, and D. F. Hunt. Expression patterns and post-translational modifications associated with mammalian histone h3 variants. J Biol Chem, 281(1):559–68, Jan 6 2006.

[137] B. L. Hamann and R. D. Blind. Nuclear phosphoinositide regulation of chromatin. Journal of Cellular Physiology, 2017.

[138] S. J. Han, F. J. DeMayo, J. Xu, S. Y. Tsai, M. J. Tsai, and B. W. O'Malley. Steroid receptor coactivator (src)-1 and src-3 differentially modulate tissue-specific activation functions of the progesterone receptor. Mol Endocrinol, 20(1):45–55, Jan 2006.

[139] P. I. Hanson, R. Roth, Y. Lin, and J. E. Heuser. Plasma membrane deformation by circular arrays of escrt-iii protein filaments. J Cell Biol, 2008.

[140] R. N. Harty, J. Paragas, M. Sudol, and P. Palese. A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with ww domains of cellular proteins: Implications for viral budding. Journal of Virology, 1999.

[141] Y. He, C. Yan, J. Fang, C. Inouye, R. Tjian, I. Ivanov, and E. Nogales. Near-atomic resolution visualization of human transcription promoter opening. Nature, 2016.

[142] W. M. Henne, N. J. Buchkovich, and S. D. Emr. The escrt pathway. Developmental Cell, 2011.

[143] W. M. Henne, N. J. Buchkovich, Y. Zhao, and S. D. Emr. The endosomal sorting complex escrt-ii mediates the assembly and architecture of escrt-iii helices. Cell, 2012.

[144] E. R. Henry, S. Bettati, J. Hofrichter, and W. A. Eaton. A tertiary two-state allosteric model for hemoglobin. Biophys Chem, 98(1-2):149–64, Jul 10 2002.

[145] O. Herkert, H. Kuhl, J. Sandow, R. Busse, and V. B. Schini-Kerth. Sex steroids used in hormonal treatment increase vascular procoagulant activity by inducing thrombin receptor (par-1) expression: role of the glucocorticoid receptor. Circulation, 104(23):2826–31, Dec 4 2001.

[146] F. Herzog, A. Kahraman, D. Boehringer, R. Mak, A. Bracher, T. Walzthoeni, A. Leitner, M. Beck, F.-U. Hartl, N. Ban, L. Malmstrom, and R. Aebersold. Structural probing of a protein phosphatase 2a network by chemical cross-linking and mass spectrometry. Science, 2012.

[147] A. Hierro, J. Sun, A. S. Rusnak, J. Kim, G. Prag, S. D. Emr, and J. H. Hurley. Structure of the escrt-ii endosomal trafficking complex. Nature, 2004.

[148] V. J. Hilser and E. Freire. Structure-based calculation of the equilibrium folding

pathway of proteins. correlation with hydrogen exchange protection factors. J Mol Biol, 262(5):756–72, Oct 11 1996.

[149] V. J. Hilser, B. García-Moreno E, T. G. Oas, G. Kapp, and S. T. Whitten. A statistical thermodynamic model of the protein ensemble. Chemical reviews, 106(5):1545–1558, 2006.

[150] V. J. Hilser and E. B. Thompson. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc Natl Acad Sci U S A, 104(20):8311–5, May 15 2007.

[151] V. J. Hilser and E. B. Thompson. Structural dynamics, intrinsic disorder, and allostery in nuclear receptors as transcription factors. J Biol Chem, 2011.

[152] S. Hirano, M. Kawasaki, H. Ura, R. Kato, C. Raiborg, H. Stenmark, and S. Wakatsuki. Double-sided ubiquitin binding of hrs-uim in endosomal protein sorting. Nat Struct and Mol Biol, 2006.

[153] A. B. Hittelman, D. Burakov, J. A. Iniguez-Lluhi, L. P. Freedman, and M. J. Garabedian. Differential regulation of glucocorticoid receptor transcriptional activation via af-1-associated proteins. EMBO J, 18(19):5380–8, Oct 1 1999.

[154] S. M. Hollenberg and R. M. Evans. Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. Cell, 1988.

[155] M. Hologne, F.-X. Cantrelle, G. Riviere, F. Guillière, X. Trivelli, and O. Walker.

Nmr reveals the interplay among the amsh sh3 binding motif, stam2, and lys63-linked diubiquitin. Journal of Molecular Biology, 2016.

[156] M. Holtzer, E. Lovett, D. d'Avignon, and A. Holtzer. Thermal unfolding in a gcn4-like leucine zipper: 13c alpha nmr chemical shifts and local unfolding curves. Biophysical journal, 1997.

[157] M. Horii, H. Shibata, R. Kobayashi, K. Katoh, C. Yorikawa, J. Yasuda, and M. Maki. Chmp7, a novel escrt-iii-related protein, associates with chmp4b and functions in the endosomal sorting pathway. Biochem J, 2006.

[158] M. Horikoshi, M. F. Carey, H. Kakidani, and R. G. Roeder. Mechanism of action of a yeast activator: Direct effect of gal4 derivatives on mammalian tfiid-promoter interactions. Cell, 1988.

[159] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Structure Function and Bioinformatics, 65(3):712–725, Nov 15 2006.

[160] P. R. Housley and W. B. Pratt. Direct demonstration of glucocorticoid receptor phosphorylation by intact l-cells. J Biol Chem, 258(7):4630–5, Apr 10 1983.

[161] P. Y. Hsu, H. K. Hsu, G. A. Singer, P. S. Yan, B. A. Rodriguez, J. C. Liu, Y. I. Weng, D. E. Deatherage, Z. Chen, J. S. Pereira, R. Lopez, J. Russo, Q. Wang,

C. A. Lamartiniere, K. P. Nephew, and T. H. Huang. Estrogen-mediated epigenetic repression of large chromosomal regions through dna looping. Genome Res, 20(6):733–44, Jun 2010.

[162] F. Huang, D. Kirkpatrick, X. Jiang, S. Gygi, and A. Sorkin. Differential regulation of egf receptor internalization and degradation by multiubiquitination within the kinase domain. Mol Cell, 2006.

[163] W. H. Hudson, C. Youn, and E. A. Ortlund. The structural basis of direct glucocorticoid-mediated transrepression. Nat Struct Mol Biol, 20(1):53–8, Jan 2013.

[164] W. H. Hudson, C. Youn, and E. A. Ortlund. Crystal structure of the mineralocorticoid receptor dna binding domain in complex with dna. PLoS One, 9(9):e107000, 2014.

[165] J. F. Hunt, A. J. Weaver, S. J. Landry, L. Gierasch, and J. Deisenhofer. The crystal structure of the GroES co-chaperonin at 2.8 angstrom resolution. Nature, 379(6560):37, 1996.

[166] J. H. Hurley. The escrt complexes. Critical Reviews in Biochemistry and Molecular Biology, 2010.

[167] M. Ikeda, E. C. Wilcox, and W. W. Chin. Different dna elements can modulate

the conformation of thyroid hormone receptor heterodimer and its transcriptional activity. J Biol Chem, 271(38):23096–104, Sep 20 1996.

[168] O. C. Ikonomov, D. Sbrissa, M. Foti, J.-L. Carpentier, and A. Shisheva. Pikfyve controls fluid phase endocytosis but not recycling/degradation of endocytosed receptors or sorting of procathepsin d by regulating multivesicular body morphogenesis. Mol Biol Cell, 2003.

[169] Y. J. Im and J. H. Hurley. Integrated structural model and membrane targeting mechanism of the human escrt-ii complex. Developmental Cell, 2008.

[170] Y. J. Im, L. Kuo, X. F. Ren, P. V. Burgos, X. Z. Zhao, F. Liu, T. R. Burke, J. S. Bonifacino, E. O. Freed, and J. H. Hurley. Crystallographic and functional analysis of the escrt-i/hiv-1 gag ptap interaction. Structure, 18(11):1536–1547, Nov 10 2010.

[171] Y. J. Im, T. Wollert, E. Boura, and J. H. Hurley. Structure and function of the escrt-ii-iii interface in multivesicular body biogenesis. Developmental Cell, 2009.

[172] R. M. Ionescu, V. F. Smith, J. C. O'Neil, and C. R. Matthews. Multistate equilibrium unfolding of Escherichia coli dihydrofolate reductase: thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. Biochemistry, 39(31):9540–9550, 2000.

[173] U. Irion and D. S. Johnston. bicoid rna localization requires specific binding of an endosomal sorting complex. Nature, 2007.

[174] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman. Zinc: a free tool to discover chemistry for biology. J Chem Inf Model, 52(7):1757–68, Jul 23 2012.

[175] R. D. Ivarie and P. H. O'Farrell. The glucocorticoid domain: steroid-mediated changes in the rate of synthesis of rat hepatoma proteins. Cell, 13(1):41–55, Jan 1978.

[176] E. K. Jaffe. Morpheeins–a new structural paradigm for allosteric regulation. Trends in biochemical sciences, 30(9):490–497, 2005.

[177] J. Jayaraman, L. Massague. Distinct oligomeric states of smad proteins in the transforming growth factor-beta pathway. J Biol Chem, 2000.

[178] I. Jelesarov and M. Lu. Thermodynamics of trimer-of-hairpins formation by the siv gp41 envelope protein. J Mol Biol, 2001.

[179] B. D. Jenkins, C. B. Pullen, and B. D. Darimont. Novel glucocorticoid receptor coactivator effector mechanisms. Trends Endocrinol Metab, 12(3):122–6, Apr 2001.

[180] C. A. P. Joazeiro, S. S. Wing, H. kuei Huang, J. D. Leverson, T. Hunter,

and Y.-C. Liu. The tyrosine kinase negative regulator c-cbl as a ring-type, e2-dependent ubiquitin-protein ligase. Science, 1999.

[181] L. Johannes, C. Wunder, and P. Bassereau. Bending ”on the rocks”–a cocktail of biophysicalmodules to build endocytic pathways. Cold Spring Harb Perspect Biol, 2014.

[182] C. Johnson, A. Cooper, and P. Stockley. Differential scanning calorimetry of thermal unfolding of the methionine repressor protein (metj) from escherichia coli. Biochemistry, 1992.

[183] C. R. Johnson and E. Freire. Structural stability of small oligomeric proteins. In D. Marshak, editor, Techniques in Protein Chemistry, volume 7, pages 459–467. 1996.

[184] C. R. Johnson, P. E. Morin, C. H. Arrowsmith, and E. Freire. Thermodynamic analysis of the structural stability of the tetrameric oligomerization domain of p53 tumor-suppressor. Biochemistry, 34(16):5309–5316, Apr 25 1995.

[185] E. E. Johnson. Physiological Role of VPS34 Phasphatidylinositol 3-Kinase in Mammalian Cells. PhD thesis, Medical College of Ohio, 2005.

[186] G. Joslyn, D. S. Richardson, R. White, and T. Alber. Dimer formation by an N-terminal coiled coil in the APC protein. Proceedings of the National Academy of Sciences, 90(23):11109–11113, 1993.

[187] J. Kach, S. D. Conzen, and R. Z. Szmulewitz. Targeting the glucocorticoid receptor in breast and prostate cancers. Sci Transl Med, 7(305):305ps19, Sep 16 2015.

[188] A. Kahraman, F. Herzog, A. Leitner, G. Rosenberger, R. Aebersold, and L. Malmstrom. Cross-link guided molecular modeling with rosetta. PLOS ONE, 2013.

[189] C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco-Jr, and C. Bustamante. The ribosome modulates nascent protein folding. Science, 2011.

[190] B. C. Kallenberger, J. D. Love, V. K. Chatterjee, and J. W. Schwabe. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. Nat Struct Biol, 10(2):136–40, Feb 2003.

[191] C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein. Structure and flexibility adaptation in nonspecific and specific protein-dna complexes. Science, 2004.

[192] T. Kamura, D. Burian, H. Khalili, S. L. Schmidt, S. Sato, W.-J. Liu, M. N. Conrad, R. C. Conaway, J. W. Conaway, and A. Shilatifard. Cloning and characterization of ell-associated proteins eap45 and eap20. J Biol Chem, 2001.

[193] S. Kangaspeska, B. Stride, R. Metivier, M. Polycarpou-Schwarz, D. Ibberson,

R. P. Carmouche, V. Benes, F. Gannon, and G. Reid. Transient cyclical methylation of promoter dna. Nature, 452(7183):112–5, Mar 6 2008.

[194] V. Karantza, A. D. Baxevanis, E. Freire, and E. N. Moudrianakis. Thermodynamic studies of the core histones: ionic strength and pH dependence of H2a-H2b dimer stability. Biochemistry, 34(17):5988–5996, 1995.

[195] V. Karantza, E. Freire, and E. N. Moudrianakis. Thermodynamic studies of the core histones: pH and ionic strength effects on the stability of the (H3-H4)/(H3- H4) 2 system. Biochemistry, 35(6):2037–2046, 1996.

[196] P. Kastner, A. Krust, B. Turcotte, U. Stropp, L. Tora, H. Gronemeyer, and P. Chambon. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms a and b. EMBO J, 9(5):1603–14, May 1990.

[197] K. Katoh, H. Shibata, K. Hatta, and M. Maki. Chmp4b is a major binding partner of the alg-2-interacting protein alix among the three chmp4 isoforms. Archives of Biochemistry and Biophysics, 2004.

[198] K. Katoh, H. Shibata, H. Suzuki, A. Nara, K. Ishidoh, E. Kominami, T. Yoshimori, and M. Maki. The alg-2-interacting protein alix associates with chmp4b, a human homologue of yeast snf7 that is involved in multivesicular body sorting. J Biol Chem, 2003.

[199] K. Katoh, H. Suzuki, Y. Terasawa, T. Mizuno, J. Yasuda, H. Shibata, and M. Maki. The penta-ef-hand protein alg-2 interacts directly with the escrt-i component tsg101, and ca2+-dependently co-localizes to aberrant endosomes with dominant-negative aaa atpase skd1/vps4b. Biochem J, 2005.

[200] D. J. Katzmann, M. Babst, and S. D. Emr. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. Cell, 106(2):145–155, 2001.

[201] S. H. Khan, S. Awasthi, C. Guo, D. Goswami, J. Ling, P. R. Griffin, J. Simons, S. S., and R. Kumar. Binding of the n-terminal region of coactivator tif2 to the intrinsically disordered af1 domain of the glucocorticoid receptor is accompanied by conformational reorganizations. J Biol Chem, 287(53):44546–60, Dec 28 2012.

[202] C. Kieffer, J. J. Skalicky, E. Morita, I. D. Domenico, D. M. Ward, J. Kaplan, and W. I. Sundquist. Two distinct modes of escrt-iii recognition are required for vps4 functions in lysosomal protein targeting and hiv-1 budding. Developmental Cell, 2008.

[203] B. Y. Kim, J. A. Olzmann, G. S. Barsh, L. S. Chin, and L. Li. Spongiform neurodegeneration-associated e3 ligase mahogunin ubiquitylates tsg101 and regulates endosomal trafficking. Mol Biol Cell, 18(4):1129–42, Apr 2007.

[204] S.-E. Kim, F. Liu, Y. J. Im, A. G. Stephen, M. J. Fivash, A. A. Waheed, E. O.

Freed, R. J. Fisher, J. H. Hurley, and T. R. Burke. Elucidation of new binding interactions with the human tsg101 protein using modified hiv-1 gag-p6 derived peptide ligands. ACS Med Chem Lett, 2011.

[205] J. M. Kinchen, K. Doukoumetzidis, J. Almendinger, L. Stergiou, A. Tosello-Trampont, C. D. Sifri, M. O. Hengartner, and K. S. Ravichandran. A pathway for phagosome maturation during engulfment of apoptotic cells. Nat Cell Biol, 2008.

[206] W. J. King and G. L. Greene. Monoclonal antibodies localize oestrogen receptor in the nuclei of target cells. Nature, 307(5953):745–7, Feb 23-29 1984.

[207] T. Kirchhausen, D. Owen, and S. C. Harrison. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. Cold Spring Harb Perspect Biol, 2014.

[208] M. L. Kirsten, R. A. Baron, M. C. Seabra, and O. Ces. Rab1a and rab5a preferentially bind to binary lipid compositions with higher stored curvature elastic energy. Molecular Membrane Biology, 2013.

[209] M. Kitano, M. Nakaya, T. Nakamura, S. Nagata, and M. Matsuda. Imaging of rab5 activity identifies essential regulators for phagosome maturation. Nature, 2008.

[210] D. Knutti, A. Kaul, and A. Kralli. A tissue-specific coactivator of steroid receptors, identified in a functional genetic screen. Mol Cell Biol, 20(7):2411–22, Apr 2000.

[211] J. Koshland, D. E., G. Nemethy, and D. Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry, 5(1):365–85, Jan 1966.

[212] M. S. Kostelansky, C. Schluter, Y. Y. Tam, S. Lee, R. Ghirlando, B. Beach, E. Conibear, and J. H. Hurley. Molecular architecture and functional model of the complete yeast escrt-i heterotetramer. Cell, 129(3):485–98, May 4 2007.

[213] M. S. Kostelansky, J. Sun, S. Lee, J. Kim, R. Ghirlando, A. Hierro, S. D. Emr, and J. H. Hurley. Structural and functional organization of the escrt-i trafficking complex. Cell, 125(1):113–26, Apr 7 2006.

[214] M. D. Krasowski, E. J. Reschly, and S. Ekins. Intrinsic disorder in nuclear hormone receptors. J Proteome Res, 7(10):4359–72, Oct 2008.

[215] R. Kumar, I. V. Baskakov, G. Srinivasan, D. W. Bolen, J. C. Lee, and E. B. Thompson. Interdomain signaling in a two-domain fragment of the human glucocorticoid receptor. J Biol Chem, 274(35):24737–41, Aug 27 1999.

[216] R. Kumar, J. C. Lee, D. W. Bolen, and E. B. Thompson. The conformation of the glucocorticoid receptor af1/tau1 domain induced by osmolyte binds co-

regulatory proteins. The Journal of biological chemistry, 276(21):18146–52, May 25 2001.

[217] R. Kumar, C. M. Moure, S. H. Khan, C. Callaway, S. L. Grimm, D. Goswami, P. R. Griffin, and D. P. Edwards. Regulation of the structurally dynamic n-terminal domain of progesterone receptor by protein-induced folding. J Biol Chem, 2013.

[218] R. Kumar and E. B. Thompson. Influence of flanking sequences on signaling between the activation function af1 and dna-binding domain of the glucocorticoid receptor. Archives of biochemistry and biophysics, 496(2):140–5, Apr 15 2010.

[219] R. Kumar, D. E. Volk, J. Li, J. C. Lee, D. G. Gorenstein, and E. B. Thompson. Tata box binding protein induces structure in the recombinant glucocorticoid receptor af1 domain. Proc Natl Acad Sci U S A, 101(47):16425–30, Nov 23 2004.

[220] I. D. Kuntz. Structure-based strategies for drug design and discovery. Science, 257(5073):1078–82, Aug 21 1992.

[221] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecule-ligand interactions. J Mol Biol, 161(2):269–88, Oct 25 1982.

[222] C. Lambert, T. Doring, and R. Prange. Hepatitis b virus maturation is sensitive to functional inhibition of escrt-iii, vps4, and 2-adaptin. Journal of Virology, 2007.

[223] A. Lange, D. Hoeller, H. Wienk, O. Marcillat, J. Lancelin, and O. Walker. Nmr reveals a different mode of binding of the stam2 vhs domain to ubiquitin and diubiquitin. Biochemistry, 2011.

[224] A. Lange, M.-B. Ismail, G. Rivière, M. Hologne, D. Lacabanne, F. Guillière, J.-M. Lancelin, I. Krimm, and O. Walker. Competitive binding of ubpy and ubiquitin to the stam2 sh3 domain revealed by nmr. FEBS Lett, 2012.

[225] C. Langelier, U. K. von Schwedler, R. D. Fisher, I. D. Domenico, P. L. White, C. P. Hill, J. Kaplan, D. Ward, and W. I. Sundquist. Human escrt-ii complex and its role in human immunodeficiency virus type 1 release. Journal of Virology, 2006.

[226] R. A. Laskowski, F. Gerick, and J. M. Thornton. The structural basis of allosteric regulation in proteins. FEBS Lett, 2009.

[227] S. Lata, M. Roessle, J. Solomons, M. Jamin, H. G. Gottlinger, D. I. Svergun, and W. Weissenhorn. Structural basis for autoinhibition of escrt-iii chmp3. J Mol Biol, 2008.

[228] S. Lata, G. Schoehn, A. Jain, R. Pires, J. Piehler, H. G. Gottlinger, and

W. Weissenhorn. Helical structures of escrt-iii are disassembled by vps4. Science, 2008.

[229] E. Lauwers, C. Jacob, and B. Andre. K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway. J Cell Biol, 2009.

[230] D. N. Lavery and I. J. McEwan. Structural characterization of the native nh2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. Biochemistry, 47(11):3360–9, Mar 18 2008.

[231] W. R. Laws and L. Brand. Analysis of 2-state excited-state reactions - fluorescence decay of 2-naphthol. Journal of Physical Chemistry, 83(7):795–802, 1979.

[232] Y. Le Drean, N. Mincheneau, P. Le Goff, and D. Michel. Potentiation of glucocorticoid receptor transcriptional activity by sumoylation. Endocrinology, 143(9):3482–9, Sep 2002.

[233] C.-P. Lee, P.-T. Liu, H.-N. Kung, M.-T. Su, H.-H. Chua, Y.-H. Chang, C.-W. Chang, C.-H. Tsai, F.-T. Liu, and M.-R. Chen. The escrt machinery is recruited by the viral bfrf1 protein to the nucleus-associated membrane for the maturation of epstein-barr virus. PLOS Pathogens, 2012.

[234] H. H. Lee, N. Elia, R. Ghirlando, J. Lippincott-Schwartz, and J. H. Hurley. Midbody targeting of the escrt machinery by a noncanonical coiled-coil in cep55. Science, 2008.

[235] J.-A. Lee, A. Beigneux, S. T. Ahmad, S. G. Young, and F.-B. Gao. Escrt-iii dysfunction causes autophagosome accumulation and neurodegeneration. Curr Biol, 2007.

[236] K. H. Lee, D. Xie, E. Freire, and L. M. Amzel. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins: Structure, Function, and Bioinformatics, 20(1):68–84, 1994.

[237] C. Lefebvre, R. Legouis, and E. Culetto. Escrt and autophagies: Endosomal functions and beyond. Semin Cell Dev Biol, 2017.

[238] J. A. Lefstin, J. R. Thomas, and K. R. Yamamoto. Influence of a steroid receptor dna-binding domain on transcriptional regulatory functions. Genes Dev, 8(23):2842–56, Dec 1 1994.

[239] M. A. Lemmon. Membrane recognition by phospholipid-binding domains. Nature Reviews Molecular Cell Biology, 2008.

[240] D. Y. Leung, Q. Hamid, A. Vottero, S. J. Szefler, W. Surs, E. Minshall, G. P. Chrousos, and D. J. Klemm. Association of glucocorticoid insensitivity with

increased expression of glucocorticoid receptor beta. Journal of Experimental Medicine, 1997.

[241] K. F. Leung, J. B. Dacks, and M. C. Field. Evolution of the multivesicular body escrt machinery; retention across the eukaryotic lineage. Traffic, 9(10):1698–716, Sep 2008.

[242] J. Li, H. N. Motlagh, C. Chakuroff, E. B. Thompson, and V. J. Hilser. Thermodynamic dissection of the intrinsically disordered n-terminal domain of human glucocorticoid receptor. The Journal of biological chemistry, 287(32):26777–87, Aug 3 2012.

[243] J. Li, J. T. White, H. Saavedra, J. O. Wrabl, H. N. Motlagh, K. Liu, J. Sowers, T. A. Schroer, E. B. Thompson, and V. J. Hilser. Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. eLife, 2017.

[244] L. Li and S. N. Cohen. Tsg101: a novel tumor susceptibility gene isolated by controlled homozygous functional knockout of allelic loci in mammalian cells. Cell, 85(3):319–29, May 3 1996.

[245] L. Li, J. Liao, J. Ruland, T. W. Mak, and S. N. Cohen. A tsg101 mdm2 regulatory loop modulates mdm2 degradation and mdm2 p53 feedback control. PNAS, 2001.

[246] W.-t. Li, R. A. Grayling, K. Sandman, S. Edmondson, J. W. Shriver, and

J. N. Reeve. Thermodynamic stability of archaeal histones. Biochemistry, 37(30):10563–10572, 1998.

[247] X. Li, S. Zhang, G. Blander, J. G. Tse, M. Krieger, and L. Guarente. Sirt1 deacetylates and positively regulates the nuclear receptor lxr. Mol Cell, 28(1):91–106, Oct 12 2007.

[248] J. M. Licata, M. Simpson-Holley, N. T. Wright, Z. Han, J. Paragas, and R. N. Harty. Overlapping motifs (ptap and ppey) within the ebola virus vp40 protein function independently as late budding domains: Involvement of host proteins tsg101 and vps-4. Journal of Virology, 2003.

[249] A.-C. Lindas, E. A. Karlsson, M. T. Lindgren, T. J. G. Ettema, and R. Bernander. A unique cell division machinery in the archaea. PNAS, 2008.

[250] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 2001.

[251] J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker. Intrinsic disorder in transcription factors. Biochemistry, 45(22):6873–6888, 2006.

[252] K. Liu, J. E. Rehfus, E. Mattson, and C. M. Kaiser. The ribosome destabilizes

native and non-native structures in a nascent multidomain protein. Protein Sci, 2017.

[253] T. Liu, D. Pantazatos, S. Li, Y. Hamuro, V. J. Hilser, and V. L. Woods. Quantitative assessment of protein structural models by comparison of H/D exchange MS data with exchange behavior accurately predicted by DXCOREX. Journal of the American Society for Mass Spectrometry, 23(1):43–56, 2012.

[254] S. W. Lockless and R. Ranganathan. Evolutionarily conserved pathways of energetic connectivity in protein families. Science, 286(5438):295–299, 1999.

[255] D. M. Lonard, R. B. Lanz, and B. W. O'Malley. Nuclear receptor coregulators and human disease. Endocr Rev, 2007.

[256] R. Loris, M. Dao-Thi, E. Bahassi, L. V. Melderen, F. Poortmans, R. Liddington, M. Couturier, and L. Wyns. Crystal structure of ccdb, a topoisomerase poison from e. coli. J Mol Biol, 1999.

[257] M. A. Loven, V. S. Likhite, I. Choi, and A. M. Nardulli. Estrogen response elements alter coactivator recruitment through allosteric modulation of estrogen receptor beta conformation. J Biol Chem, 276(48):45282–8, Nov 30 2001.

[258] C. Lu, L.-Z. Mi, M. J. Grey, J. Zhu, E. Graef, S. Yokoyama, and T. A. Springer. Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor. Mol Cell Biol, 2010.

[259] M. Lu, M. Stoller, S. Wang, J. Liu, M. Fagan, and J. Nunberg. Structural and functional analysis of interhelical interactions in the human immunodeficiency virus type 1 gp41 envelope glycoprotein by alanine-scanning mutagenesis. J Virol, 2001.

[260] N. Z. Lu and J. A. Cidlowski. Translational regulatory mechanisms generate n-terminal glucocorticoid receptor isoforms with unique transcriptional target genes. Molecular cell, 18(3):331–42, Apr 29 2005.

[261] N. Z. Lu, J. B. Collins, S. F. Grissom, and J. A. Cidlowski. Selective regulation of bone cell apoptosis by translational isoforms of the glucocorticoid receptor. Mol Cell Biol, 2007.

[262] Q. Lu, L. W. Hope, M. Brasch, C. Reinhard, and S. N. Cohen. Tsg101 interaction with hrs mediates endosomal trafficking and receptor down-regulation. PNAS, 2003.

[263] B. F. Luisi, W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, and P. B. Sigler. Crystallographic analysis of the interaction of the glucocorticoid receptor with dna. Nature, 352(6335):497–505, Aug 8 1991.

[264] M. L. Lupher-Jr, N. Rao, N. L. Lill, C. E. Andoniou, S. Miyake, E. A. Clark, B. Druker, and H. Band. Cbl-mediated negative regulation of the syk tyrosine kinase. J Biol Chem, 1998.

[265] M. Mani, C. Chen, V. Amblee, H. Liu, T. Mathur, G. Zwicke, S. Zabad, B. Patel, J. Thakkar, and C. J. Jeffery. Moonprot: a database for proteins that are known to moonlight. Nucleic Acids Res, 2015.

[266] A. P. Manuel, J. Lambert, and M. T. Woodside. Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. PNAS, 2015.

[267] M. C. Marchetti, B. Di Marco, G. Cifone, G. Migliorati, and C. Riccardi. Dexamethasone-induced apoptosis of thymocytes: role of glucocorticoid receptor-associated src kinase and caspase-8 activation. Blood, 101(2):585–93, Jan 15 2003.

[268] J. Martin-Serrano, A. Yarovoy, D. Perez-Caballero, and P. D. Bieniasz. Divergent retroviral late-budding domains recruit vacuolar protein sorting factors by using alternative adaptor proteins. PNAS, 2003.

[269] E. Martinez and W. Wahli. Cooperative binding of estrogen receptor to imperfect estrogen-responsive dna elements correlates with their synergistic hormone-dependent enhancer activity. EMBO J, 8(12):3781–91, Dec 1 1989.

[270] J. R. Mayers, I. Fyfe, A. L. Schuh, E. R. Chapman, J. M. Edwardson, and A. Audhya. Escrt-0 assembles as a heterotetrameric complex on membranes and binds multiple ubiquitinylated cargoes simultaneously. J Biol Chem, 2011.

[271] S. Mayor, R. G. Parton, and J. G. Donaldson. Clathrin-independent pathways of endocytosis. Cold Spring Harb Perspect Biol, 2014.

[272] J. McCullough, M. J. Clague, and S. Urbe. Amsh is an endosome-associated ubiquitin isopeptidase. JCB, 2004.

[273] J. McCullough, A. K. Clippinger, N. Talledge, M. L. Skowyra, M. G. Saunders, T. V. Naismith, L. A. Colf, P. Afonine, C. Arthur, W. I. Sundquist, P. I. Hanson, and A. Frost. Structure and membrane remodeling activity of escrt-iii helical polymers. Science, 2015.

[274] J. McCullough, R. D. Fisher, F. G. Whitby, W. I. Sundquist, and C. P. Hill. Alix-chmp4 interactions in the human escrt pathway. PNAS, 2008.

[275] P. O. McGowan, A. Sasaki, A. C. D'Alessio, S. Dymov, B. Labonte, M. Szyf, G. Turecki, and M. J. Meaney. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci, 12(3):342–8, Mar 2009.

[276] A. N. McKeown, J. T. Bridgham, D. W. Anderson, M. N. Murphy, E. A. Ortlund, and J. W. Thornton. Evolution of dna specificity in a transcription factor family produced a new gene regulatory module. Cell, 159(1):58–68, Sep 25 2014.

[277] G. Medina, Y. Zhang, Y. Tang, E. Gottwein, M. L. Vana, F. Bouamr, J. Leis,

and C. A. Carter. The functionally exchangeable l domains in rsv and hiv-1 gag direct particle release through pathways linked by tsg101. Traffic, 2005.

[278] S. H. Meijsing, M. A. Pufall, A. Y. So, D. L. Bates, L. Chen, and K. R. Yamamoto. Dna binding site sequence directs glucocorticoid receptor structure and activity. Science, 324(5925):407–10, Apr 17 2009.

[279] D. Meinhold, S. Boswell, and W. Colon. P61a mutation in the factor for inversion stimulation results in a thermostable dimeric intermediate. Biochemistry, 2005.

[280] C. J. Merrifield and M. Kaksonen. Endocytic accessory factors and regulation of clathrin-mediated endocytosis. Cold Spring Harb Perspect Biol, 2014.

[281] M. E. Meyer, A. Pornon, J. W. Ji, M. T. Bocquel, P. Chambon, and H. Gronemeyer. Agonistic and antagonistic activities of ru486 on the functions of the human progesterone receptor. EMBO J, 9(12):3923–32, Dec 1990.

[282] P. J. Mitchell and R. Tjian. Transcriptional regulation in mamnalian cells by sequence-specific dna binding proteins. Science, 1989.

[283] E. Mizuno, T. Iura, A. Mukai, T. Yoshimori, N. Kitamura, and M. Komada. Regulation of epidermal growth factor receptor down-regulation by ubpy-mediated deubiquitination at endosomes. Mol Biol Cell, 2005.

[284] E. Mizuno, K. Kawahata, M. Kato, N. Kitamura, and M. Komada. Stam proteins

bind ubiquitinated proteins on the early endosome via the vhs domain and ubiquitin-interacting motif. Mol Biol Cell, 2003.

[285] K. H. Moberg, S. Schelble, S. K. Burdick, and I. K. Hariharan. Mutations in erupted, the drosophila ortholog of mammalian tumor susceptibility gene 101, elicit non-cell-autonomous overgrowth. Developmental Cell, 2005.

[286] J. Monod, J. Wyman, and J. P. Changeux. On the nature of allosteric transitions: A plausible model. J Mol Biol, 12:88–118, May 1965.

[287] E. Morita, V. Sandrin, S. L. Alam, D. M. Eckert, S. P. Gygi, and W. I. Sundquist. Identification of human MVB12 proteins as ESCRT-I subunits that function in HIV budding. Cell host & microbe, 2(1):41–53, 2007.

[288] E. Morita and W. I. Sundquist. Retrovirus budding. Annu. Rev. Cell Dev. Biol., 20:395–425, 2004.

[289] H. N. Motlagh and V. J. Hilser. Agonism/antagonism switching in allosteric ensembles. Proc Natl Acad Sci U S A, 109(11):4134–9, Mar 13 2012.

[290] H. N. Motlagh, D. Toptygin, C. M. Kaiser, and V. J. Hilser. Single-Molecule Chemo-Mechanical Spectroscopy Provides Structural Identity of Folding Intermediates. Biophysical journal, 110(6):1280–1290, 2016.

[291] H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser. The ensemble nature of allostery. Nature, 2014.

[292] J. F. Mouillet, V. C. Henrich, M. Lezzi, and M. Vogtli. Differential control of gene activity by isoforms a, b1 and b2 of the drosophila ecdysone receptor. Eur J Biochem, 268(6):1811–9, Mar 2001.

[293] R. Muromoto, K. Sugiyama, A. Takachi, S. Imoto, N. Sato, T. Yamamoto, K. Oritani, K. Shimoda, and T. Matsuda. Physical and functional interactions between daxx and dna methyltransferase 1-associated protein, dmap1. J Immunol, 172(5):2985–93, Mar 1 2004.

[294] K. P. Murphy, V. Bhakuni, D. Xie, and E. Freire. Molecular basis of co-operativity in protein folding: III. Structural identification of cooperative folding units and folding intermediates. Journal of molecular biology, 227(1):293–306, 1992.

[295] K. P. Murphy, D. Xie, K. S. Thompson, L. M. Amzel, and E. Freire. Entropy in biological binding processes: estimation of translational entropy loss. Proteins: Structure, Function, and Bioinformatics, 18(1):63–67, 1994.

[296] T. Muziol, E. Pineda-Molina, R. B. Ravelli, A. Zamborlini, Y. Usami, H. Gottlinger, and W. Weissenhorn. Structural basis for budding by the escrt-iii factor chmp3. Developmental Cell, 2006.

[297] S. Nagashima, M. Takahashi, S. Jirintai, T. Tanaka, T. Nishizawa, J. Yasuda, and H. Okamoto. Tumour susceptibility gene 101 and the vacuolar protein

sorting pathway are required for the release of hepatitis e virions. Journal of General Virology, 2011.

[298] V. Nahoum, E. Perez, P. Germain, F. Rodriguez-Barrios, F. Manzo, S. Kammerer, G. Lemaire, O. Hirsch, C. A. Royer, H. Gronemeyer, A. R. de Lera, and W. Bourguet. Modulators of the structural dynamics of the retinoid x receptor to reveal receptor function. Proc Natl Acad Sci U S A, 104(44):17323–8, Oct 30 2007.

[299] M. J. G. Naima Ismaili, Raymond Blind. Stabilization of the unliganded glucocorticoid receptor by tsg101. Journal of Biological Chemistry, 2005.

[300] Z. Nie, Y. Xue, D. Yang, S. Zhou, B. J. Deroo, T. K. Archer, and W. Wang. A specificity and targeting subunit of a human swi/snf family-related chromatin-remodeling complex. Mol Cell Biol, 20(23):8879–88, Dec 2000.

[301] E. Nikko, A.-M. Marini, and B. Andre. Permease recycling and ubiquitination status reveal a particular role for bro1 in the multivesicular body pathway. J Biol Chem, 2003.

[302] D. B. Nikolov, H. Chen, E. D. Halay, A. Hoffman, R. G. Roeder, and S. K. Burley. Crystal structure of a human tata box-binding protein/tata element complex. PNAS, 1996.

[303] C. Nilsson, U. Johansson, A.-C. Johansson, K. K\a agedal, and K. Öllinger.

Cytosolic acidification and lysosomal alkalinization during TNF-α induced apoptosis in U937 cells. Apoptosis, 11(7):1149–1159, 2006.

[304] M. Nocula-Ługowska, G. Rymarczyk, M. Lisowski, and A. Ożyhar. Isoform-specific variation in the intrinsic disorder of the ecdysteroid receptor n-terminal domain. Proteins: Structure, Function, and Bioinformatics, 76(2):291–308, 2009.

[305] A. W. Norman, M. T. Mizwicki, and D. P. Norman. Steroid-hormone rapid actions, membrane receptors and a conformational ensemble model. Nat Rev Drug Discov, 3(1):27–41, Jan 2004.

[306] M. G. Oakley and P. S. Kim. A buried polar interaction can direct the relative orientation of helices in a coiled coil. Biochemistry, 37(36):12603–12610, 1998.

[307] H. Pan, J. C. Lee, and V. J. Hilser. Binding sites in escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. Proc Natl Acad Sci U S A, 2000.

[308] V. G. Pandey, S. Jain, A. Rana, N. Puri, S. K. Arudra, B. Mopidevi, M. Kaw, A. Nasjletti, and A. Kumar. Dexamethasone promotes hypertension by allele-specific regulation of the human angiotensinogen gene. J Biol Chem, Jan 7 2015.

[309] V. G. Panse, C. P. Swaminathan, J. J. Aloor, A. Surolia, and R. Varadarajan. Unfolding thermodynamics of the tetrameric chaperone, SecB. Biochemistry, 39(9):2362–2369, 2000.

[310] G. D. Paolo and P. D. Camilli. Phosphoinositides in cell regulation and membrane dynamics. Nature, 2006.

[311] N. Pashkova, L. Gakhar, S. C. Winistorfer, A. B. Sunshine, M. Rich, M. J. Dunham, L. Yu, and R. C. Piper. The yeast alix homolog bro1 functions as a ubiquitin receptor for protein sorting into multivesicular endosomes. Developmental Cell, 2013.

[312] S. Patnaik, S. P. Georgea, E. Phama, S. Roya, K. Singha, J. M. Mariadasonb, and S. Khurana. By moonlighting in the nucleus, villin regulates epithelial plasticity. Mol Biol Cell, 2016.

[313] M. F. Perutz. Stereochemistry of cooperative effects in haemoglobin: haem-haem interaction and the problem of allostery. Nature, 1970.

[314] M. F. Perutz, G. Fermi, B. Luisi, B. Shaanan, and R. C. Liddington. Stereochemistry of cooperative mechanisms in hemoglobin. Cold Spring Harb Symp Quant Biol, 52:555–65, 1987.

[315] M. F. Perutz, A. J. Wilkinson, M. Paoli, and G. G. Dodson. The stereochemical

mechanism of the cooperative effects in hemoglobin revisited. Annual review of biophysics and biomolecular structure, 27(1):1–34, 1998.

[316] C. M. Petit, J. Zhang, P. J. Sapienza, E. J. Fuentes, and A. L. Lee. Hidden dynamic allostery in a PDZ domain. Proceedings of the National Academy of Sciences, 106(43):18249–18254, 2009.

[317] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. Ucsf chimera–a visualization system for exploratory research and analysis. J Comput Chem, 25(13):1605–12, Oct 2004.

[318] J. W. Pike and N. M. Sleator. Hormone-dependent phosphorylation of the 1,25-dihydroxyvitamin d3 receptor in mouse fibroblasts. Biochem Biophys Res Commun, 131(1):378–85, Aug 30 1985.

[319] I. Pilecka, M. Banach-Orlowska, and M. Miaczynska. Nuclear functions of endocytic proteins. European Journal of Cell Biology, 2007.

[320] E. Pineda-Molina, H. Belrhali, A. J. Piefer, I. Akula, P. Bates, and W. Weissenhorn. The crystal structure of the c-terminal domain of vps28 reveals a conserved surface required for vps20 recruitment. Traffic, 2006.

[321] R. Pires, B. Hartlieb, L. Signor, G. Schoehn, S. Lata, M. Roessle, C. Moriscot, S. Popov, A. Hinz, M. Jamin, V. Boyer, R. Sadoul, E. Forest, D. I. Svergun,

H. G. Gottlinger, and W. Weissenhorn. A crescent-shaped alix dimer targets escrt-iii chmp4 filaments. Structure, 2009.

[322] S. Ponnusamy, C. C. Coss, T. Thiyagarajan, K. Watts, D.-J. Hwang, Y. He, L. A. Selth, I. J. McEwan, C. B. Duke, J. Pagadala, G. Singh, R. W. Wake, C. Ledbetter, W. D. Tilley, T. Moldoveanu, J. T. Dalton, D. D. Miller, and R. Narayanan. Novel selective agents for the degradation of androgen receptor variants to treat castration-resistant prostate cancer. Cancer Res, 2017.

[323] O. Pornillos, S. L. Alam, D. R. Davis, and W. I. Sundquist. Structure of the tsg101 uev domain in complex with the ptap motif of the hiv-1 p6 protein. Nat Struct Biol, 9(11):812–7, Nov 2002.

[324] O. Pornillos, D. S. Higginson, K. M. Stray, R. D. Fisher, J. E. Garrus, M. Payne, G. P. He, H. E. Wang, S. G. Morham, and W. I. Sundquist. Hiv gag mimics the tsg101-recruiting activity of the human hrs protein. J Cell Biol, 162(3):425–34, Aug 4 2003.

[325] H. Poukka, U. Karvonen, O. A. Janne, and J. J. Palvimo. Covalent modification of the androgen receptor by small ubiquitin-like modifier 1 (sumo-1). Proc Natl Acad Sci U S A, 97(26):14145–50, Dec 19 2000.

[326] P. L. Privalov, N. N. Khechinashvili, and B. P. Atanasov. Thermodynamic analysis of thermal transitions in globular proteins. i. calorimetric study of chymotrypsinogen, ribonuclease and myoglobin. Biopolymers, 1971.

[327] L. K. Putney and D. L. Barber. Na-H exchange-dependent increase in intracellular pH times G2/M entry and transition. Journal of Biological Chemistry, 278(45):44645–44649, 2003.

[328] B. Pyrzynska, I. Pilecka, and M. Miaczynska. Endocytic proteins in the regulation of nuclear signaling, transcription and tumorigenesis. Molecular Oncology, 2009.

[329] D. Pérez-Sala, D. Collado-Escobar, and F. Mollinedo. Intracellular alkalinization suppresses lovastatin-induced apoptosis in HL-60 cells through the inactivation of a pH-dependent endonuclease. Journal of Biological Chemistry, 270(11):6235–6242, 1995.

[330] M. Raab, M. Gentili, H. de Belly, H. R. Thiam, P. Vargas, A. J. Jimenez, F. Lautenschlaeger, R. Voituriez, A. M. Lennon-Dumenil, N. Manel, and M. Piel. Escrt iii repairs nuclear envelope ruptures during cell migration to limit dna damage and cell death. Science, 2016.

[331] J. Rafferty, W. Somers, I. Saint-Girons, and S. Phillips. Three-dimensional crystal structures of escherichia coli met repressor with and without corepressor. Nature, 1989.

[332] C. Raiborg, K. G. Bache, D. J. Gillooly, I. H. Madshus, E. Stang, and H. Stenmark. Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes. Nat Cell Biol, 2002.

[333] C. Raiborg, K. G. Bache, A. Mehlum, E. Stang, and H. Stenmark. Hrs recruits clathrin to early endosomes. EMBO J, 2001.

[334] G. Raposo and W. Stoorvogel. Extracellular vesicles: Exosomes, microvesicles, and friends. J Cell Biol, 2013.

[335] F. J. RauscherIII, D. R. Cohen, T. Curran, T. J.Bos, P. K. Vogt, D. Bohmann, R. Tjian, and B. R. FranzaJr. Fos-associated protein p39 is the product of the jun proto-oncogene. Science, 1988.

[336] F. Reggiori and H. R. Pelham. Sorting of proteins into multivesicular bodies: ubiquitin-dependent and -independent targeting. EMBO J, 2001.

[337] S. E. Reichheld, Z. Yu, and A. R. Davidson. The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. Proceedings of the National Academy of Sciences, 106(52):22263–22268, 2009.

[338] J. Reid, S. M. Kelly, K. Watt, N. C. Price, and I. J. McEwan. Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. influence of structure-stabilizing solutes and protein-protein interactions. J Biol Chem, 277(22):20079–86, May 31 2002.

[339] X. Ren and J. H. Hurley. Vhs domains of escrt-0 cooperate in high-avidity binding to polyubiquitinated cargo. EMBO J, 2010.

[340] X. Ren, D. P. Kloer, Y. C. Kim, R. Ghirlando, L. F. Saidi, G. Hummer, and J. H. Hurley. Hybrid structural model of the complete human escrt-0 complex. Structure, 2009.

[341] J. P. Renaud, N. Rochel, M. Ruff, V. Vivat, P. Chambon, H. Gronemeyer, and D. Moras. Crystal structure of the rar-gamma ligand-binding domain bound to all-trans retinoic acid. Nature, 378(6558):681–9, Dec 14 1995.

[342] E. Rieser, S. M. Cordier, and H. Walczak. Linear ubiquitination: a newly discovered regulator of cell signalling. Trends in Biochemical Sciences, 2013.

[343] J. S. Robinson, D. J. Klionsky, L. M. Banta, and S. D. Emr. Protein sorting in saccharomyces cerevisiae: isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases. Mol Cell Biol, 8(11):4936–48, Nov 1988.

[344] S. C. Roemer, D. C. Donham, L. Sherman, V. H. Pon, D. P. Edwards, and M. E. Churchill. Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. Mol Endocrinol, 20(12):3042–52, Dec 2006.

[345] K. Ronacher, K. Hadley, C. Avenant, E. Stubsrud, J. Simons, S. S., A. Louw, and J. P. Hapgood. Ligand-selective transactivation and transrepression via the glucocorticoid receptor: role of cofactor interaction. Mol Cell Endocrinol, 299(2):219–31, Feb 27 2009.

[346] J. H. Rothman, I. Howald, and T. H. Stevens. Characterization of genes required for protein sorting and vacuolar function in the yeast saccharomyces cerevisiae. EMBO J, 8(7):2057–65, Jul 1989.

[347] B. Roucourt, S. Meeussen, J. Bao, P. Zimmermann, and G. David. Heparanase activates the syndecan-syntenin-alix exosome pathway. Cell Research, 2015.

[348] N. Roudier, C. Lefebvre, and R. Legouis. Cevps-27 is an endosomal protein required for the molting and the endocytic trafficking of the low- density lipoprotein receptor-related protein 1 in caenorhabditis elegans. Traffic, 2005.

[349] P. E. Row, H. Liu, S. Hayes, R. Welchman, P. Charalabous, K. Hofmann, M. J. Clague, C. M. Sanders, and S. Urbe. The mit domain of ubpy constitutes a chmp binding and endosomal localization signal required for efficient epidermal growth factor receptor degradation. J Biol Chem, 2007.

[350] H. Russcher, V. A. Dalm, F. H. de Jong, A. O. Brinkmann, L. J. Hofland, S. W. Lamberts, and J. W. Koper. Associations between promoter usage and alternative splicing of the glucocorticoid receptor gene. J Mol Endocrinol, 38(1-2):91–8, Feb 2007.

[351] H. Russcher, E. F. van Rossum, F. H. de Jong, A. O. Brinkmann, S. W. Lamberts, and J. W. Koper. Increased expression of the glucocorticoid

receptor-a translational isoform as a result of the er22/23ek polymorphism. Mol Endocrinol, 19(7):1687–96, Jul 2005.

[352] A. P. Sagona, I. P. Nezis, N. M. Pedersen, K. Liestol, J. Poulton, T. E. Rusten, R. I. Skotheim, C. Raiborg, and H. Stenmark. Ptdins(3)p controls cytokinesis through kif13a-mediated recruitment of fyve-cent to the midbody. Nat Cell Biol, 2010.

[353] D. D. Sakai, S. Helms, J. Carlstedt-Duke, J. A. Gustafsson, F. M. Rottman, and K. R. Yamamoto. Hormone-mediated repression: a negative glucocorticoid response element from the bovine prolactin gene. Genes Dev, 2(9):1144–54, Sep 1988.

[354] R. M. Sapolsky, L. M. Romero, and A. U. Munck. How do glucocorticoids influence stress responses? integrating permissive, suppressive, stimulatory, and preparative actions. Endocr Rev, 21(1):55–89, Feb 2000.

[355] D. Sbrissa, O. C. Ikonomov, and A. Shisheva. Pikfyve, a mammalian ortholog of yeast fab1p lipid kinase, synthesizes 5-phosphoinositides: Effect of insulin. JBC, 1999.

[356] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. Nat Methods, 2012.

[357] J. Schindelin, C. T. Rueden, M. C. Hiner, and K. W. Eliceiri. The imagej ecosystem: An open platform for biomedical image analysis. Molecular Reproduction and Development, 2015.

[358] A. E. Schmidt, T. Miller, S. L. Schmidt, R. Shiekhattar, and A. Shilatifard. Cloning and characterization of the eap30 subunit of the ell complex that confers derepression of transcription by rna polymerase ii. J Biol Chem, 1999.

[359] W. T. Schrader and B. W. O'Malley. Progesterone-binding components of chick oviduct. iv. characterization of purified subunits. J Biol Chem, 247(1):51–9, Jan 10 1972.

[360] A. L. Schuh, M. Hanna, K. Quinney, L. Wang, A. Sarkeshik, J. R. Y. III, and A. Audhya. The vps-20 subunit of the endosomal sorting complex escrt-iii exhibits an open conformation in the absence of upstream activation. Biochem J, 2015.

[361] J. W. Schwabe, L. Chapman, J. T. Finch, and D. Rhodes. The crystal structure of the estrogen receptor dna-binding domain bound to dna: how receptors discriminate between their response elements. Cell, 75(3):567–78, Nov 5 1993.

[362] J. W. Schwabe, L. Chapman, and D. Rhodes. The oestrogen receptor rec-

ognizes an imperfectly palindromic response element through an alternative side-chain conformation. Structure, 3(2):201–13, Feb 15 1995.

[363] J. W. Schwabe, D. Neuhaus, and D. Rhodes. Solution structure of the dna-binding domain of the oestrogen receptor. Nature, 348(6300):458–61, Nov 29 1990.

[364] S. Sen, S. L. Deshmane, R. Kaminski, S. Amini, and P. K. Datta. Non-metabolic role of pkm2 in regulation of the hiv-1 ltr. Journal of Cellular Physiology, 2017.

[365] P. L. Shaffer, A. Jivan, D. E. Dollins, F. Claessens, and D. T. Gewirth. Structural basis of androgen receptor binding to selective androgen response elements. Proc Natl Acad Sci U S A, 101(14):4758–63, Apr 6 2004.

[366] Y. Shang and M. Brown. Molecular determinants for the tissue specificity of serms. Science, 295(5564):2465–8, Mar 29 2002.

[367] E. K. Shea, J. D. Klemm, and others. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science, 254(5031):539, 1991.

[368] E. K. Shea, R. Rutkowski, and P. S. Kim. Evidence that the leucine zipper is a coiled coil. Science, 243(4890):538, 1989.

[369] Q.-T. Shen, A. L. Schuh, Y. Zheng, K. Quinney, L. Wang, M. Hanna, J. C. Mitchell, M. S. Otegui, P. Ahlquist, Q. Cui, and A. Audhya. Structural analysis

and modeling reveals new mechanisms governing escrt-iii spiral filament assembly. J Cell Biol, 2014.

[370] A. Shilatifard. Identification and purification of the holo-ell complex. J Biol Chem, 1998.

[371] H.-W. Shin, M. Hayashi, S. Christoforidis, S. Lacas-Gervais, S. Hoepfner, M. R. Wenk, J. Modregger, S. Uttenweiler-Joseph, M. Wilm, A. Nystuen, W. N. Frankel, M. Solimen, P. D. Camill, and M. Zerial. An enzymatic cascade of rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. J Cell Biol, 2005.

[372] W. Shu, J. Liu, H. Ji, and M. Lu. Core structure of the outer membrane lipoprotein from escherichia coli at 1.9 a resolution. J Mol Biol, 2000.

[373] J. Skog, T. Wurdinger, S. van Rijn, D. H. Meijer, L. Gainche, M. Sena-Esteves, W. T. Curry-Jr, B. S. Carter, A. M. Krichevsky, and X. O. Breakefield. Glioblastoma microvesicles transport rna and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol, 2008.

[374] T. Slagsvold, R. Aasland, S. Hirano, K. G. Bache, C. Raiborg, D. Trambaiolo, S. Wakatsuki, and H. Stenmark. Eap45 in mammalian escrt-ii binds ubiquitin via a phosphoinositide-interacting glue domain. J Biol Chem, 2005.

[375] F. R. Smith and G. K. Ackers. Experimental resolution of cooperative free

energies for the ten ligation states of human hemoglobin. Proceedings of the National Academy of Sciences, 82(16):5347–5351, 1985.

[376] W. T. Sneada, C. C. Haydena, A. K. Gadoka, C. Zhaoa, E. M. Laferb, P. Rangamanic, and J. C. Stachowiak. Membrane fission by protein crowding. PNAS, 2017.

[377] J. Solomons, C. Sabin, E. Poudevigne, Y. Usami, D. L. Hulsik, P. Macheboeuf, B. Hartlieb, H. Gottlinger, and W. Weissenhorn. Structural basis for escrt-iii chmp3 recruitment of amsh. Structure, 2011.

[378] S. Spinelli, Q. Z. Liu, P. M. Alzari, P. H. Hirel, and R. J. Poljak. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie, 73(11):1391–1396, 1991.

[379] R. S. Spolar and J. Record, M. T. Coupling of local folding to site-specific binding of proteins to dna. Science, 263(5148):777–84, Feb 11 1994.

[380] J. H. Stack, P. K. Herman, P. V. Schu, and S. D. Emr. A membrane-associated complex containing the vps15 protein kinase and the vps34 pi 3-kinase is essential for protein sorting to the yeast lysosome-like vacuole. EMBO J, 1993.

[381] R. V. Stahelin, F. Long, K. Diraviyam, K. S. Bruzik, D. Murray, and W. Cho.

Phosphatidylinositol 3-phosphate induces the membrane penetration of the fyve domains of vps27p and hrs. J Biol Chem, 2002.

[382] J. Stamos, M. X. Sliwkowski, and C. Eigenbrot. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J Biol Chem, 2002.

[383] D. B. Starr, W. Matsui, J. R. Thomas, and K. R. Yamamoto. Intracellular receptors use a common mechanism to interpret signaling information at response elements. Genes Dev, 10(10):1271–83, May 15 1996.

[384] P. W. Staskus and W. C. Johnson Jr. Conformational transition of hyaluronic acid in aqueous-organic solvent monitored by vacuum ultraviolet circular dichroism. Biochemistry, 27(5):1522–1527, 1988.

[385] D. R. Stauffer, T. L. Howard, T. Nyun, and S. M. Hollenberg. Chmp1 is a novel nuclear matrix protein affecting chromatin structure and cell-cycle progression. J Cell Science, 2001.

[386] C. Steif, P. Weber, H. J. Hinz, J. Flossdorf, G. Cesareni, and M. Kokkinidis. Subunit interactions provide a significant contribution to the stability of the dimeric four-. alpha.-helical-bundle protein ROP. Biochemistry, 32(15):3867–3876, 1993.

[387] Y. Stijf-Bultsma, L. Sommer, M. Tauber, M. Baalbaki, P. Giardoglou, D. R.

Jones, K. A. Gelato, J. van Pelt, Z. Shah, H. Rahnamoun, C. Toma, K. E. Anderson, P. Hawkins, S. M. Lauberth, A.-P. G. Haramis, D. Hart, W. Fischle, and N. Divecha. The basal transcription complex component taf3 transduces changes in nuclear phosphoinositides into transcriptional output. Mol Cell, 2015.

[388] B. Strack, A. Calistri, S. Craig, E. Popova, and H. G. Gottlinger. Aip1/alix is a binding partner for hiv-1 p6 and eiav p9 functioning in virus budding. Cell, 2003.

[389] S. J. Stray, P. Ceres, and A. Zlotnick. Zinc ions trigger conformational change and oligomerization of hepatitis b virus capsid protein. Biochemistry, 2004.

[390] D. Suckau, A. Resemann, M. Schuerenberg, P. Hufnagel, J. Franzen, and A. Holle. A novel maldi lift-tof/tof mass spectrometer for proteomics. Anal Bioanal Chem, 376(7):952–65, Aug 2003.

[391] S. Sun, L. Li1, F. Yang, X. Wang, F. Fan, M. Yang, C. Chen, X. Li, H.-W. Wang, and S.-F. Sui. Cryo-em structures of the atp-bound vps4-e233q hexamer and its complex with vta1 at near-atomic resolution. Nature Communications, 2017.

[392] Z. Sun, J. Pan, W. X. Hope, S. N. Cohen, and S. P. Balk. Tumor susceptibility gene 101 protein represses androgen receptor transactivation and interacts with p300. Cancer, 86(4):689–96, Aug 15 1999.

[393] W. I. Sundquist, H. L. Schubert, B. N. Kelly, G. C. Hill, J. M. Holton, and C. P. Hill. Ubiquitin recognition by the human TSG101 protein. Molecular cell, 13(6):783–789, 2004.

[394] S. Suzuki, K. Tamai, M. Watanabe, M. Kyuuma, M. Ono, K. Sugamura, and N. Tanaka. Amsh is required to degrade ubiquitinated proteins in the central nervous system. Biochem Biophys Res Commun, 2011.

[395] D. Szapary, Y. Huang, and J. Simons, S. S. Opposing effects of corepressor and coactivators in determining the dose-response curve of agonists, and residual agonist activity of antagonists, for glucocorticoid receptor-regulated gene expression. Mol Endocrinol, 13(12):2108–21, Dec 1999.

[396] D. Szapary, M. Xu, and J. Simons, S. S. Induction properties of a transiently transfected glucocorticoid-responsive gene vary with glucocorticoid receptor concentration. J Biol Chem, 271(48):30576–82, Nov 29 1996.

[397] G. M. Süel, S. W. Lockless, M. A. Wall, and R. Ranganathan. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nature structural & molecular biology, 10(1):59–69, 2003.

[398] H. Takahashi, J. R. Mayers, L. Wang, J. M. Edwardson, and A. Audhya. Hrs and stam function synergistically to bind ubiquitin-modified cargoes in vitro. Biophysical journal, 2015.

[399] W. S. Talbot, E. A. Swyryd, and D. S. Hogness. Drosophila tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. Cell, 73(7):1323–37, Jul 2 1993.

[400] H. Teo, O. Perisic, B. Gonzalez, and R. L. Williams. Escrt-ii, an endosome-associated complex required for protein sorting: Crystal structure and interactions with escrt-iii and membranes. Developmental Cell, 2004.

[401] T. C. Terwilliger and D. Eisenberg. The structure of melittin. I. Structure determination and partial refinement. Journal of Biological Chemistry, 257(11):6010–6015, 1982.

[402] D. Thain, T. Tannenbaum, and M. Livny. Distributed computing in practice: The condor experience. Concurrency and Computation: Practice and Experience, 17(2-4):323–356, 2005.

[403] E. B. Thompson and R. Kumar. Dna binding of nuclear hormone receptors influences their structure and function. Biochem Biophys Res Commun, 306(1):1–4, Jun 20 2003.

[404] K. S. Thompson, C. R. Vinson, and E. Freire. Thermodynamic characterization of the structural stability of the coiled-coil region of the bzip transcription factor gcn4. Biochemistry, 32(21):5491–5496, Jun 1 1993.

[405] S. G. Thota, C. P. Unnikannan, S. R. Thampatty, R. Manorama, and R. Bhan-

dari. Inositol pyrophosphates regulate rna polymerase i-mediated rrna transcription in saccharomyces cerevisiae. Biochem J, 2015.

[406] P. J. Thul, L. Akesson, M. Wiking, D. Mahdessian, A. Geladaki, H. A. Blal, T. Alm, A. Asplund, L. Bjork, L. M. B. andAnna Backstrom, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, A. Sivertsson, E. Sjostedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K. S. Lilley, M. Uhlén, and E. Lundberg. A subcellular map of the human proteome. Science, 2017.

[407] R. Thulasi, D. V.Harbour, and E. Thompson. Suppression of c-myc is a critical step in glucocorticoid-induced human leukemic cell lysis. J Biol Chem, 1993.

[408] M. M. Tirado, C. L. Martinez, and J. G. Delatorre. Comparison of theories for the translational and rotational diffusion-coefficients of rod-like macromolecules - application to short dna fragments. Journal of Chemical Physics, 81(4):2047–2052, 1984.

[409] M. J. Todd, N. Semo, and E. Freire. The structural stability of the HIV-1 protease. Journal of molecular biology, 283(2):475–488, 1998.

[410] P. Tompa. Unstructural biology coming of age. Current opinion in structural biology, 21(3):419–425, 2011.

[411] P. Tontonoz, E. Hu, R. A. Graves, A. I. Budavari, and B. M. Spiegelman. mppar gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev, 8(10):1224–34, May 15 1994.

[412] D. Toptygin. Analysis of time-dependent red shifts in fluorescence emission from tryptophan residues in proteins. Fluorescence Spectroscopy and Microscopy: Methods and Protocols, pages 215–256, 2014.

[413] D. Toptygin, R. S. Savtchenko, N. D. Meadow, and L. Brand. Homogeneous spectrally- and time-resolved fluorescence emission from single-tryptophan mutants of iia(glc) protein. Journal of Physical Chemistry B, 105(10):2043–2055, Mar 15 2001.

[414] K. Trajkovic, C. Hsu, S. Chiantia, L. Rajendran, D. Wenzel, F. Wieland, P. Schwille, B. Brugger, and M. Simons. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. Science, 2008.

[415] J. W. Truman, W. S. Talbot, S. E. Fahrbach, and D. S. Hogness. Ecdysone receptor expression in the cns correlates with stage-specific responses to ecdysteroids during drosophila and manduca development. Development, 120(1):219–34, Jan 1994.

[416] B. Ulmschneider, B. K. Grillo-Hill, M. Benitez, D. R. Azimova, D. L. Barber, and T. G. Nystul. Increased intracellular pH is necessary for adult epithelial and embryonic stem cell differentiation. J Cell Biol, 215(3):345–355, 2016.

[417] V. N. Uversky, C. J. Oldfield, and A. K. Dunker. Showing your id: intrinsic disorder as an id for recognition, regulation and cell signaling. J Mol Recognit, 18(5):343–84, Sep-Oct 2005.

[418] H. Valadi, K. Ekstrom, A. Bossios, M. Sjostrand, J. J. Lee, and J. O. Lotvall. Exosome-mediated transfer of mrnas and micrornas is a novel mechanism of genetic exchange between cells. Nat Cell Biol, 2007.

[419] S. Vardhana, K. Choudhuri, R. Varma, and M. L. Dustin. Essential role of ubiquitin and tsg101 protein in formation and function of the central supramolecular activation cluster. Immunity, 2010.

[420] J. Vertrees, P. Barritt, S. Whitten, and V. J. Hilser. COREX/BEST server: a web browser-based program that calculates regional stability variations within protein structures. Bioinformatics, 21(15):3318–3319, 2005.

[421] J. Virgin, C. E., T. P. Ha, D. R. Packan, G. C. Tombaugh, S. H. Yang, H. C. Horner, and R. M. Sapolsky. Glucocorticoids inhibit glucose transport and glutamate uptake in hippocampal astrocytes: implications for glucocorticoid neurotoxicity. J Neurochem, 57(4):1422–8, Oct 1991.

[422] U. K. von Schwedler, M. Stuchell, B. Muller, D. M. Ward, H.-Y. Chung, E. Morita, H. E. Wang, T. Davis, G.-P. He, D. M. Cimbora, A. Scott, H.-G. Krausslich, J. Kaplan, S. G. Morham, and W. I. Sundquist. The protein network of hiv budding. Cell, 2003.

[423] R. L. Wagner, J. W. Apriletti, M. E. McGrath, B. L. West, J. D. Baxter, and R. J. Fletterick. A structural role for hormone in the thyroid hormone receptor. Nature, 378(6558):690–7, Dec 14 1995.

[424] A. D. Wallace and J. A. Cidlowski. Proteasome-mediated glucocorticoid receptor degradation restricts transcriptional signaling by glucocorticoids. J Biol Chem, 276(46):42714–21, Nov 16 2001.

[425] A. E. Wallberg, K. E. Neely, J.-A. Gustafsson, J. L. Workman, A. P. H. Wright, and P. A. Grant. Histone acetyltransferase complexes can mediate transcriptional activation by the major glucocorticoid receptor activation domain. Mol Cell Biol, 1999.

[426] K. A. Walters, U. Simanainen, and D. J. Handelsman. Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models. Hum Reprod Update, 16(5):543–58, Sep-Oct 2010.

[427] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model, 25(2):247–60, Oct 2006.

[428] J. M. Wang and P. A. Kollman. Automatic parameterization of force field by systematic search and genetic algorithms. Journal of Computational Chemistry, 22(12):1219–1228, Sep 2001.

[429] W. Wang and C. J. Jeffery. An analysis of surface proteomics results reveals novel candidates for intracellular/surface moonlighting proteins in bacteria. Mol BioSyst, 2016.

[430] A. Warnmark, A. Wikstrom, A. P. Wright, J. A. Gustafsson, and T. Hard. The n-terminal regions of estrogen receptor alpha and beta are unstructured in vitro and show different tbp binding properties. J Biol Chem, 276(49):45939–44, Dec 7 2001.

[431] M. Watanabe, Y. Yanagi, Y. Masuhiro, T. Yano, H. Yoshikawa, J. unn Yanagisawa, and S. Kato. A putative tumor suppressor, tsg101, acts as a transcriptional suppressor through its coiled-coil domain. Biochem Biophys Res Commun, 2008.

[432] M. Watanabe, Y. Yanagi, Y. Masuhiro, T. Yano, H. Yoshikawa, J. Yanagisawa, and S. Kato. A putative tumor suppressor, tsg101, acts as a transcriptional suppressor through its coiled-coil domain. Biochem Biophys Res Commun, 245(3):900–5, Apr 28 1998.

[433] A. L. Watters, P. Deka, C. Corrent, D. Callender, G. Varani, T. Sosnick, and D. Baker. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. Cell, 128(3):613–24, Feb 9 2007.

[434] B. A. Webb, M. Chimenti, M. P. Jacobson, and D. L. Barber. Dysregulated pH: a

perfect storm for cancer progression. Nature Reviews Cancer, 11(9):671–677, 2011.

[435] J. C. Webster, C. M. Jewell, J. E. Bodwell, A. Munck, M. Sar, and J. A. Cidlowski. Mouse glucocorticoid receptor phosphorylation status influences multiple functions of the receptor protein. J Biol Chem, 272(14):9287–93, Apr 4 1997.

[436] W. V. Welshons, M. E. Lieberman, and J. Gorski. Nuclear localization of unoccupied oestrogen receptors. Nature, 307(5953):747–9, Feb 23-29 1984.

[437] J. T. White, H. N. Motlagh, J. Li, E. B. Thompson, and V. J. Hilser. Nuclear Receptors: From Structure to the Clinic Chapter: Allosteric Regulation and Intrinsic Disorder in Nuclear Hormone Receptors. Springer, 2015.

[438] J. T. White, D. Toptygin, R. Cohen, N. Murphy, and V. J. Hilser. Structural stability of the coiled-coil domain of tumor susceptibility gene (tsg)-101. Biochemistry, 2017.

[439] P. Whitley, B. J. Reaves, M. Hashimoto, A. M. Riley, B. V. L. Potter, and G. D. Holman. Identification of mammalian vps24p as an effector of phosphatidylinositol 3,5-bisphosphate-dependent endosome compartmentalization. J Biol Chem, 2003.

[440] S. T. Whitten, E. B. Garcia-Moreno, and V. J. Hilser. Local conformational

fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. Proc Natl Acad Sci U S A, 2005.

[441] B. Winkel. When an enzyme isn't just an enzyme anymore. Journal of Experimental Botany, 2017.

[442] D. M. Wolf and V. C. Jordan. Gynecologic complications associated with long-term adjuvant tamoxifen therapy for breast cancer. Gynecol Oncol, 45(2):118–28, May 1992.

[443] J. R. Wood, G. L. Greene, and A. M. Nardulli. Estrogen response elements function as allosteric modulators of estrogen receptor conformation. Mol Cell Biol, 18(4):1927–34, Apr 1998.

[444] J. R. Wood, V. S. Likhite, M. A. Loven, and A. M. Nardulli. Allosteric modulation of estrogen receptor conformation by different estrogen response elements. Mol Endocrinol, 15(7):1114–26, Jul 2001.

[445] J. Wyman and S. Gill. Binding and linkage: functional chemistry of biological macromolecules. 1990.

[446] J. Xiao, H. Xia, J. Zhou, I. F. Azmi, B. A. Davies, D. J. Katzmann, and Z. Xu. Structural basis of vta1 function in the multivesicular body sorting pathway. Developmental Cell, 2008.

[447] D. Xie and E. Freire. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins: Structure, Function, and Bioinformatics, 19(4):291–301, 1994.

[448] D. Xie and E. Freire. Structure based prediction of protein folding intermediates. Journal of molecular biology, 242(1):62–80, 1994.

[449] W. Xie, L. Li, and S. Cohen. Cell cycle-dependent subcellular localization of the tsg101 protein and mitotic and nuclear abnormalities associated with tsg101 deficiency. Proc Natl Acad Sci U S A, 1998.

[450] K. R. Yamamoto and B. M. Alberts. In vitro conversion of estradiol-receptor protein to its nuclear form: dependence on hormone and dna. Proc Natl Acad Sci U S A, 69(8):2105–9, Aug 1972.

[451] Y. Yan, E. Winograd, A. Viel, T. Cronin, S. Harrison, and D. Branton. Crystal structure of the repetitive segments of spectrin. Science, 1993.

[452] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, and C. T. Walsh. Genetically encoded short peptide tag for versatile protein labeling by sfp phosphopantetheinyl transferase. PNAS, 2005.

[453] C. Yorikawa, H. Shibata, S. Waguri, K. Hatta, M. Horii, K. Katoh, T. Kobayashi,

Y. Uchiyama, and M. Maki. Human chmp6, a myristoylated escrt-iii protein, interacts directly with an escrt-ii component eap20 and regulates endosomal cargo sorting. Biochem J, 2005.

[454] M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, and G. Dollinger. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. PNAS, 2000.

[455] H. Yuan, S. Wang, W. Yang, S. Finkel, and R. Johnson. The structure of fis mutant pro61ala illustrates that the kink within the long alpha-helix is not due to the presence of the proline residue. J Biol Chem, 1994.

[456] M. Zen, M. Canova, C. Campana, S. Bettio, L. Nalotto, M. Rampudda, R. Ramonda, L. Iaccarino, and A. Doria. The kaleidoscope of glucorticoid effects on immune system. Autoimmun Rev, 10(6):305–10, Apr 2011.

[457] L. Zhang, M. Xu, E. Scotti, Z. J. Chen, and P. Tontonoz. Both k63 and k48 ubiquitin linkages signal lysosomal degradation of the ldl receptor. Journal of Lipid Research, 2013.

[458] X. Zhang, J. Gureasko, K. Shen, P. Cole, and J. Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell, 2006.

[459] Q. Zhong, Y. Chen, D. Jones, and W.-H. Lee. Perturbation of tsg101 protein affects cell cycle progression. Cancer Research, 1998.

[460] H. Zhu, Z. Liang, and G. Li. Rabex-5 is a rab22 effector and mediates a rab22-rab5 signaling cascade in endocytosis. Mol Biol Cell, 2009.

Vita

Biographical Sketch

- Son of a home-ec teacher and a cowpoke turned nuclear engineer. Born and raised in the city of Chesapeake in the Old Dominion (1989).

My undergraduate thesis was largely to do with the endocrinology of mice. I was interested in why two populations of mice had one divergent characteristic: one reproduced year round, the other didn't. Prof Heideman and I looked at the obvious suspects—testosterone, estrogen, corticosterone (a glucocorticoid), and many others. There were no differences between the populations. It emphasized to me that hormone levels are not a ready source of evolution. Except

when released locally, hormones will affect the entire body, so a small increase in testosterone could throw everything into disarray. Instead, it would be more advantageous to change things in the target organ: Express more or different receptors.

When I first tread onto the Hopkins campus, these thoughts were in my head. I was broadly interested in several aspects of the biology department. Everything from RNA riboswitches to endocytosis. Yet, I had an aching desire to better understand hormone receptor biology. In particular, how could expression of different receptor isoforms affect the biological output? Only by immersing myself in the biophysics of hormone biology could I hope to understand how the molecular events lead to cellular and organismal events. So with no understanding of biophysics, immerse myself I did. I've found quantitative knowledge a rewarding experience.

Previous Work Experience:

- Undergraduate volunteer in the lab of Paul Heideman, College of William & Mary (2007–2011)
- Research student in the lab of Tim Bartness, Georgia State University (Summer 2010)
- Graduate studies in the lab of Vince Hilser, Johns Hopkins (2012 – current)

- Teaching assistant for:
  - Undergraduate Biochemistry (Fall 2012)
  - Undergraduate Cell Biology (Spring 2013)
  - Graduate Biophysical Chemistry (Spring and Fall of 2014, Head TA for Fall 2015)

Degrees:

- B.S. Biology cum laude (with mathematics minor and senior honors). College of William & Mary. Thesis entitled: "Endocrine Variation in a Population of *Peromyscus leucopus* with Divergent Life-History Strategies." (May 2011; GPA 3.50 overall, 3.70 Biology)

Scholarships and Awards:

- Charles Center Domestic Scholarship for Summer Research (2009)
- College of William & Mary's Dean's list (Fall 2008 and 2009)
- Owen Scholars Fellowship (Johns Hopkins University, 2011)
- Dupont Teaching Award in Biochemistry (Fall 2012)

Grants:

- Howard Hughes Medical Institute Grant to the College of William & Mary, Freshman Research Funding (2007-2008)
- Howard Hughes Medical Institute Grant to the College of William & Mary, Undergraduate Science Education and Research Funding (2009-2010)
- Student Conference Travel Award, College of William & Mary Undergraduate Science Education and Research Program, Funded by a Howard Hughes Medical Institute grant to the College of William & Mary

Presentations:

- Fall Undergraduate Science Symposium 2009: “Hormones of Peromyscus leucopus”
- Spring Undergraduate Science Symposium 2010: “Hormones of Peromyscus leucopus”
- Poster Presentation at American Society of Mammalogists 2010 Meeting in Laramie, Wyoming: “Life history strategies and endocrine variation: genetic variation in reproduction, food intake, body composition, glucose, and feeding-related hormones in a population of Peromyscus leucopus”
- Poster Presentation at JHU Biology retreat 2013: “Investigation of a protein cofactor of glucocorticoid receptor”
- Speaker and Poster Presenter at Protein Folding Consortium 2016: “Characterization of the coiled-coil tetramer of Tumor Susceptibility Gene-101”
- Speaker and Poster Presenter at JHU Biology retreat 2016: “The interaction of Tumor Susceptibility Gene 101 and Glucocorticoid Receptor”

Original Research:

- White, JT; DeSanto, CL; Gibbons, C; Lardner, CK; Panakos, A; Rais, S; Sharp, K.; Sullivan, SD; Tidhar, W; Wright, L; Berrigan, D; Heideman, P. (2014) “Insulins, Leptin and Feeding in a Population of *Peromyscus leucopus* (white-footed mouse) with Variable Fertility.” Hormones and Behavior.
- White, JT; Toptygin, D; Cohen, R; Murphy, N; Hilser, VJ. (2017) “Structural Stability of the Coiled-Coil Domain of Tumor Susceptibility Gene (TSG)-101.” Biochemistry
- Li, J.; White, JT; Saavedra, H; Wrabl, J. O.; Motlagh, H. N.; Liu, K.; Sowers, J.; Schroer, T. A.; Thompson E. B.; Hilser, V. J. (2017) “Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor” eLife

Reviews/Book Chapters:

- White, JT; Motlagh, HN; Li, J; Thompson, EB; Hilser, VJ. (2015) “Allosteric Regulation and Intrinsic Disorder in Nuclear Hormone Receptors.” (eds.) McEwan, IJ and Kumar, R. Nuclear Receptors: From Structure to Clinic. Part1, Chapter 4.